NISTIR 8275

Mission Critical Voice Quality of Experience Access Time Measurement Methods

Jaden Pieper
Jesse Frey
Chelsea Greene
Zainab Soetan
Tim Thompson
Stephen Voran
Donald Bradshaw

NIST

**National Institute of
Standards and Technology**
U.S. Department of Commerce

NISTIR 8275

Mission Critical Voice Quality of Experience Access Time Measurement Methods

Jaden Pieper
Jesse Frey
Chelsea Greene
Zainab Soetan
Tim Thompson
Donald Bradshaw
Communications Technology Laboratory
Stephen Voran
Institute for Telecommunication Sciences NTIA

October 2019

U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Undersecretary of Commerce for Standards and Technology

Disclaimer

Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Acknowledgments

Abstract

Access time generally describes the time associated with the establishment of a talk path upon user request to speak and has been identified as a key component of quality of experience (QoE) in communications. NIST's Public Safety Communications Research (PSCR) Division developed a method to measure and quantify the access time of any push-to-talk (PTT) communication system. This measurement method is a follow-on development to the mouth-to-ear (M2E) latency measurement system presented in Ref. [1]. Here, a broad definition of access time is created that is applicable across multiple PTT technologies.

In this paper, a speech intelligibility-based access delay measurement system is introduced. This system measures the Modified Rhyme Test (MRT) intelligibility of a target word based on when PTT was pushed within a predefined message. It relies only on speech going into and coming out of a voice communications system and PTT timing, so it functions as a fair platform to compare different technologies. Example measurements were performed across the following land mobile radio (LMR) technologies: analog direct and conventional modes, and digital Project 25 (P25) direct, trunked Phase 1, and trunked Phase 2 modes.

Key words

Table of Contents

List of Tables

List of Figures

Acronyms

3GPP 3rd Generation Partnership Project. 5–7, 10

ABC-MRT Articulation Band Correlation Modified Rhyme Test. i, 9–11, 21, 22, 29, 30

D2D Device-to-Device. i, 3, 5

FM frequency modulation. 2

KPI key performance indicator. i, 3, 7

LMR land mobile radio. i, 2–6, 20, 30

LTE Long Term Evolution. 2, 3, 5, 29, 30

M2E mouth-to-ear. i, 1, 3, 6, 10, 12, 17, 22, 27, 29, 34, 35

MCPTT Mission Critical Push-to-Talk. i, 5, 7

MCV mission critical voice. i, 1, 2

MRT Modified Rhyme Test. i, 1, 8, 9, 11, 12, 21, 30

P25 Project 25. i, 2, 6, 20, 29

ProSe Proximity Services. i, 3, 5

PSCR Public Safety Communications Research. i, 2, 8

PTT push-to-talk. i, 1–3, 5, 6, 10–14, 16–22, 24, 25, 27, 29, 30, 32–35

QoE quality of experience. i, 1–3, 19, 29

QoS quality of service. i, 1, 2, 6, 7

RMSE root mean square error. 21

SUT system under test. 1, 9, 16, 21, 30

TIA Telecommunications Industry Association. 6, 7, 10

UE user equipment. 3, 5–7

Symbols

α Intelligibility scaling factor. 10, 14, 15, 20, 23

I_0 Asymptotic intelligibility. 10, 13–15, 23, 24, 29

λ Logistic parameter, intelligibility curve steepness. 15, 20, 23–25, 27, 29

L_w Word length. 13, 14

P_1 First utterance of MRT keyword. 12–14, 21, 22, 29

P_2 Second utterance of MRT keyword. 12–14, 21, 22, 27, 29

T Time preceding P_1 and P_2 in audio clips. 11–14, 18

t Word invariant time. 14

t_0 Logistic parameter, intelligibility curve midpoint. 15, 23–25, 27, 29

τ_A Access delay, function of α. 14, 15, 25, 27

T_{ptt} Time PTT pressed within an audio clip. 14, 15

V_{Rx} Receive volume. 19, 20, 33

V_{Tx} Transmit volume. 19, 20, 33

1. Introduction

This paper builds on the mouth-to-ear (M2E) latency measurements system of Ref. [1] in developing mission critical voice (MCV) quality of experience (QoE) measurements for public safety communications systems. This paper specifically demonstrates a technology-independent measurement system for access delay. This paper, combined with Ref. [1], demonstrates end-to-end access time measurements of push-to-talk (PTT) systems.

QoE measurements differ from quality of service (QoS) measurements that are classically used to describe system performance. Rather than focusing on internal system-specific measurements, QoE measurements focus on the external events that describe user interaction with the system. Users interact with PTT communications systems by pressing the PTT button, speaking into the system, or listening to speech coming out of the system. Hence, any QoE-based measurement considers audio going into and coming out of the communications system, as well as the timing of the PTT button presses.

This paper defines a few terms related to access time. End-to-end access time describes the total amount of time from a button push by a transmitting user until a receiving user hears intelligible speech. This time is composed of two sources: M2E latency and access delay. M2E latency is defined as the time between speech input into one device and its output from another (with this definition microphone-to-speaker latency is technically more correct, but not conventionally used). Access delay characterizes the time it takes for a system to handle potential channel allocation upon a PTT request, as well as the time for the communications devices to turn on all required systems to transmit and receive voice.

From a user experience point of view, access delay can be defined as the minimum length of time a user must wait between pressing a PTT button on a communications device and starting to speak to ensure that the start of the message is not lost. The emphasis in the access delay definition on whether or not the start of the message is lost implies the need for audio intelligibility to be directly tied to measurement of access. Throughout this paper the word "intelligibility" specifically means Modified Rhyme Test (MRT) intelligibility.

The value of access delay has tangible operational consequences. If a transmitting user presses the PTT button, waits for the appropriate amount of time (the access delay), and then begins speaking, the receiving user will hear the entirety of the spoken message. If the transmitting user waits for a time that is less than the access delay, the receiving user will not hear the start of the message. This can result in a request for re-transmission (slowing down communications and operations) or a misunderstanding (with consequences ranging from insignificant to fatal). Thus, it is most efficient and safest for the transmitting user to wait for the full access delay, but no longer, before starting to speak.

2. Research Summary

The access delay measurement system presented here is an aggregate measurement system, where results are only achieved after a multitude of tests. Multiple tests are required in order to capture variations in speech intelligibility and variations in how the system under

1

test (SUT) responds when the PTT button is pressed. Due to the relationship between access delay and intelligibility, it was not initially feasible to develop a system with a one-to-one relationship between pressing a PTT button and measuring the access time of the system. Instead the measurement system relies on varying when a PTT button is pressed across a series of trials and measuring the intelligibility of the first word of a message. The measurement system is primarily designed as a lab measurement with a controlled radio frequency (RF) environment, due to the sensitivity of intelligibility scores to RF conditions.

The measurement is independent of any signaling the system might give a user to indicate when access is granted. In practice, users would not speak before receiving notice that access was granted. In order to measure the minimum amount of wait time required to transmit intelligible audio, it is necessary to force the communications system to incrementally cut off speech. This mimics a user speaking before receiving an access grant notification. The measurement system is designed so that, over the course of the test, there will be a transmission that represents a realistic use case where the transmitting user waits the appropriate amount of time before speaking. Essentially, this is what the measurement system is designed to find.

The end result of an access delay measurement is a curve that describes the amount of time it takes the system to achieve a fractional portion of its asymptotic intelligibility. Here asymptotic intelligibility describes the intelligibility of the system when given more than ample time to prepare for sending and receiving voice communications.

Example measurements were performed using analog frequency modulation (FM) and Project 25 (P25) radio systems. P25 measurements were done in both direct mode and using a Phase 1 and Phase 2 trunked system. Analog measurements were done in direct mode and using a conventional analog repeater. Measurements were also performed using a simple radio system replacement, referred to as the PTT gate. The PTT gate is an electronic switch in the audio path that will pass or not pass audio based on the PTT signal. The PTT gate describes the lower limit for access delay values for this measurement system, as it functions as the simplest possible PTT communications system.

Researchers are encouraged to replicate the end-to-end access time measurement system developed by Public Safety Communications Research (PSCR). Data gathered during the measurement experiments are available at https://doi.org/10.18434/M32083 and the MATLAB code developed for the measurement system is available at https://doi.org/10.18434/M32085. Limited technical assistance is available by contacting Tim Thompson at (303) 497-6613 or at tim.thompson@nist.gov.

3. Background

The National Institute of Standards and Technology's (NIST) PSCR group held a roundtable event in March 2017 with industry and public safety representatives to identify expectations and metrics that would enable PSCR to understand, measure, monitor, and predict MCV QoE across land mobile radio (LMR), Long Term Evolution (LTE), and future technologies. Quantifying QoE is a departure from traditional QoS metrics which focus on

network and device performance. QoE focuses on the end users and their experience with the communications system. The four key performance indicators (KPIs) identified in the roundtable event were M2E latency, end-to-end access time, audio quality/intelligibility, and access probability/retention. Reference [1] addressed the measurement of M2E latency. In this report, access delay, a component of end-to-end access time, is quantified by leveraging intelligibility measurement capabilities. The measurement system presented here is based on an audio in/audio out approach, meaning that measurements are done by feeding audio into the system and processing the audio that comes out. This results in a device-independent system that allows for direct measurement comparisons across a variety of communications technologies. The remaining KPIs will be addressed in future work.

Access delay describes the minimum amount of time a user must wait between pressing PTT and starting to speak to ensure that the beginning of the message is not lost. This definition directly relates to the QoE for users. Users must wait at least the access delay of a system before speaking for their message to be transmitted and understood by the receiving user. Systems with highly variable access delay may force users to be more cognizant of their communications systems, rather than be focused on the job at hand. Access time measurements are generally associated with communications systems where channels must be allocated upon PTT requests. Such systems typically inform users of when they can speak by providing some sort of signal that they have been granted access. However, not all PTT communications systems provide such signaling, which is detailed further in Sec. 3.1. Even when systems do provide such feedback to the transmitting user, there is no guarantee that the system is actually fully setup to transmit intelligible speech. Thus, intelligibility lends itself as a useful metric for making a platform for fair access delay measurements across technologies, as it relates directly to the impact access delay can have on the user experience.

3.1 Communications Technology Overview

The end-to-end access time measurement system can be used with any PTT voice communications system. It was developed primarily with LMR and LTE PTT technologies in mind and provides a platform for fair performance comparisons. For any wireless communications technology, when a user presses a PTT button to speak, the user's voice is transmitted over some RF channel. For an LMR system, which channel a user speaks over and how that is determined depends on if the system is operating in direct, conventional, or trunked mode. For LTE, the subscribers can either operate on-network or off-network. In the on-network scenario, the subscriber, or user equipment (UE), uses specific resource blocks (15 kHz subcarriers) allocated by the eNodeB base station to transmit and receive the voice packets, which is comparable to trunked mode operation in an LMR system. In the off-network scenario, the UE works much more like LMR direct mode and is referred to as Proximity Services (ProSe) or Device-to-Device (D2D) communications.

3.1.1 LMR Systems

LMR systems have three primary operating modes: direct, conventional, and trunked. Direct mode communications require only the transmitting and receiving devices, while both conventional and trunked modes use repeaters as relays to extend the range of the communications system. Diagrams of all three modes of LMR communications can be seen in Fig. 1.

(a) Direct mode

(b) Conventional mode

(c) Trunked mode
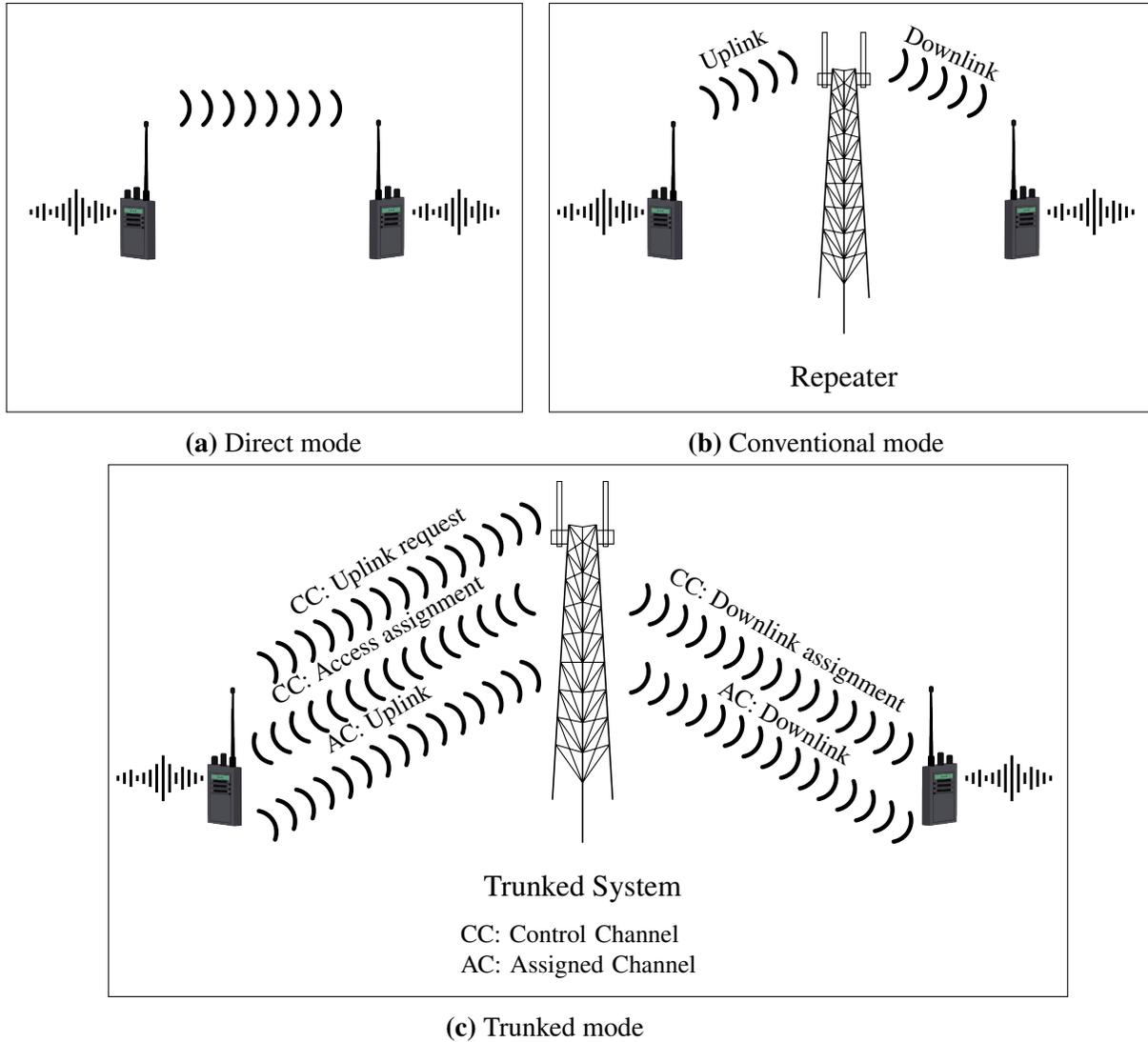
Fig. 1. LMR system diagrams

In direct mode communications, the transmitting device communicates directly with the receiving devices. Therefore, only one frequency is required for communications. In conventional mode, the transmitting device sends its message to a repeater which then broadcasts the message out to receiving devices. This requires two frequencies: the uplink

frequency for communications from transmitting devices to the repeater and the downlink frequency for communications from the repeater to receiving devices.

Trunked mode allows for multiple talk groups, or separate groups of users, to share a smaller number of frequency channels. It does this by allocating a pair of uplink and downlink frequencies for a talk group only upon an access request. Here a transmitting user presses a PTT button, which signals the repeater with a request for access. The repeater then allocates a channel pair for that talk group and broadcasts that information out to all users. The transmitting user is notified that access has been granted and can then send a message over the allocated uplink frequency. That message is then relayed to other users by the repeater over the allocated downlink frequency. Trunked communications therefore require one additional frequency pair for the control channel, so that devices can communicate with the repeater for channel requests.

Both direct and conventional modes lack a concept of floor control, an assurance that users have been granted a channel and that other users are hearing them. For these technologies, the only indication that a channel has been granted, or is already in use, is that speech is coming out of a receive radio. Generally, this lack of floor control confirmation works in direct and conventional modes because there is little system setup in these modes; however, it can be an issue when two users press PTT at nearly the same moment. In that situation, receiving users may only hear one of the speakers, their signals may interfere, and the two speakers would have no idea that an issue was occurring.

In trunked mode, the system signals transmitting users when they have been granted floor control. This is necessary due to the variable channel setup time inherent in trunked communications. There is also potential in trunked systems for no channels to be available for a requesting user, in which case the system must signal to the user that there is not a channel available. The manner in which the system signals the transmitting user about channel status can vary, but a simple beep is commonly used.

3.1.2   LTE Systems

Mission Critical Push-to-Talk (MCPTT) LTE systems have two primary modes of operation: on-network and off-network. Figure 2 shows both on-network and off-network LTE scenarios. The off-network mode is similar to LMR direct mode in that the UEs communicate directly with each other without communicating through the eNodeB infrastructure. The commonly referenced D2D communication in 3rd Generation Partnership Project (3GPP) standards is ProSe. ProSe uses some of the resource blocks from the uplink LTE frequencies; this channel is referred to as the sidelink. The over-the-air interface between the UEs is defined as the PC5 interface. Although ProSe has been defined and ratified in the 3GPP LTE standards, no manufacturer has developed a ProSe chipset for UEs as of the writing of this report.

For on-network operation, MCPTT LTE operates much like LMR trunked mode. For example, multiple talkgroups are allowed, and group and private calls are supported in much the same way as LMR trunked mode. To initiate a call, the MCPTT LTE UE makes a

5

(a) Off-network MCPTT utilizing the ProSe (D2D) Sidelink PC5 interface

CC: Control Channel, TC: Traffic Channel

(b) On-network MCPTT utilizing the LTE user interface
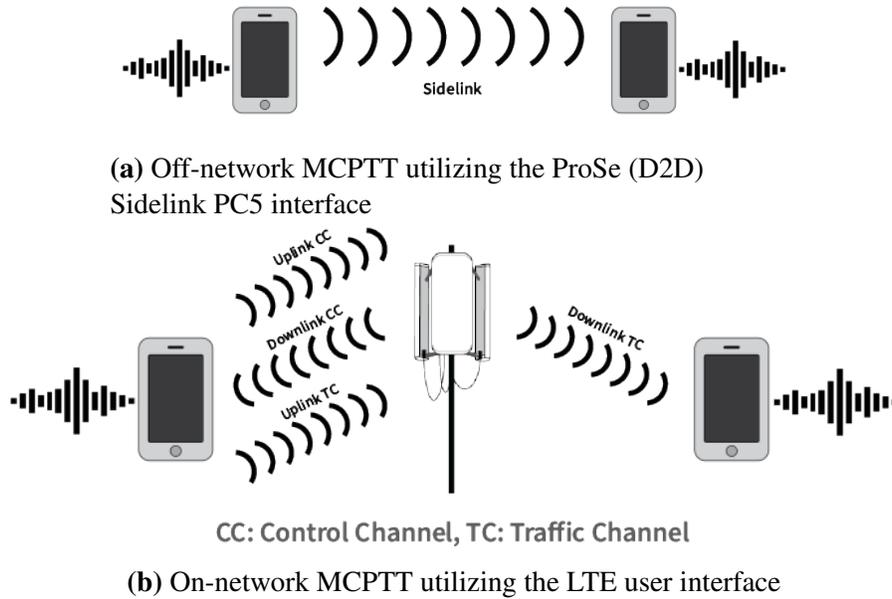
Fig. 2. (a) LTE off-network and (b) on-network diagrams

channel request by the user pushing a PTT button. The eNodeB allocates certain resource blocks on both the uplink and downlink to support the PTT call. The transmitting UE receives acknowledgment that the call setup is complete, and the user can begin speaking. The 3GPP standards also allow acknowledgment from the receiving devices. Group calls can be set up with or without acknowledgments from receiving users [2].

3.2   Mouth-to-ear Latency

M2E latency characterizes the time between speech being input into one device and its output from another. It is not specific to PTT technologies, but is a fundamental delay component of all communications systems. Furthermore, definitions for M2E latency are fairly consistent across standards such as 3GPP and the Telecommunications Industry Association (TIA), which establishes the P25 standards. Reference [1] contains more detailed information on M2E latency.

3.3   Access Time in Other Standards

Access time is defined differently across technology standards. This is emblematic of the differences in how channels are granted and how communications occur in general across different technologies. Furthermore, access time definitions have classically been QoS based and have potential to misrepresent what system users might experience.

In LMR communications, access time is generally associated with trunked mode and the idea of a user issuing a channel request and waiting for the system to grant that request. This description of access time is much more aligned with a conventional QoS perspective.

6

The TIA defines access time as "the time required from the initiation of a user push-to-talk until a traffic channel is granted, and transmission on that channel has begun. This measurement includes the time for both the subscriber equipment and the RF subsystem equipment" [3]. This definition focuses completely on the output power of the transmitting radio as can be seen in Fig. 3. This definition does not distinguish between transmitting RF power while audio processing and vocoding are still starting up and transmitting actual intelligible voice. This definition also does not account for any setup time that might be added by the receiving radio or any additional infrastructure.
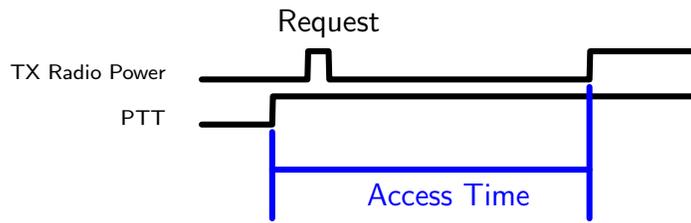
Fig. 3. TIA standard for access time

3GPP defines MCPTT access time as "the time between when an MCPTT User requests to speak (normally by pressing the MCPTT control on the MCPTT UE) and when this user gets a signal to start speaking" [2]. This is shown as KPI 1 in Fig. 4. They also define end-to-end access time as "the time between when an MCPTT User requests to speak (normally by pressing the MCPTT control on the MCPTT UE) and when this user gets a signal to start speaking, including call establishment (if applicable), and possibly acknowledgment from first receiving user before voice can be transmitted" [2]. Again, both of these definitions are QoS based; they describe access time in terms of system signaling and device acknowledgment. While the end-to-end access time definition accounts for receiving devices acknowledging the channel grant, they do not account for the crux of a user experience: how long must the talker wait in order to ensure that intelligible voice is received by another user.
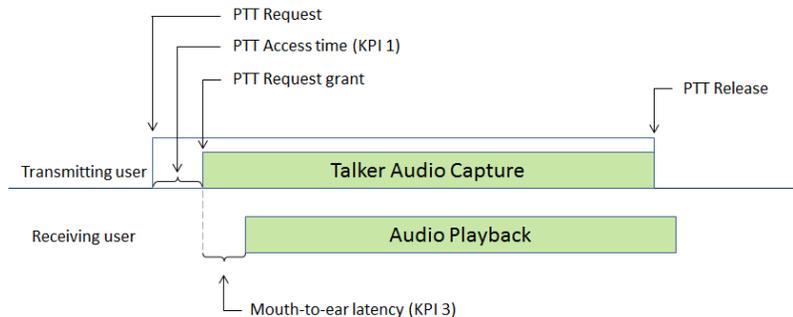
Fig. 4. Illustration of MCPTT access time and M2E latency[1]

---

[1]3GPP Technical Specification (TS) 22.179 MCPTT over LTE Stage 1 (Release 14)

3.4   Speech Intelligibility

3.4.1   Modified Rhyme Test

PSCR measures speech intelligibility using the MRT. This selection was made after extensive discussions with numerous public safety partners when PSCR was formulating the project described in Ref. [4]. The selection stems from the protocol described in Ref. [5] for testing "face-to-face" voice communication through the mask associated with a self-contained breathing apparatus in noisy environments. This protocol specifies a testing environment and then specifies the MRT as the actual testing mechanism. The MRT is fully defined in Ref. [6]. In an MRT trial, a subject must identify the word presented from a set of six words that rhyme. At the conclusion of the MRT, every condition under test receives an intelligibility rating that is based on the fraction of words correctly identified.

The MRT protocol specifies 50 sets containing 6 words (or keywords) each. Some of the sets contain words that rhyme in the strict sense, for example, "bed," "led," "fed," "red," "wed," and "shed." Other sets contain words that rhyme in a more general sense—the words display some type of phonetic similarity. An example is the set "dug," "dung," "duck," "dud," "dub," and "dun." In the MRT, each keyword is presented after a carrier phrase: "Please select the word —." For example, when the test word is "bed," the entire sentence is "Please select the word bed." Two female and two male talkers were used to record the MRT keywords in these sentences as described in Ref. [7].

In the MRT, a subject hears a sentence (e.g., "Please select the word bed") and then performs that task (selecting the proper word) using a graphical user interface. There are always six words to choose from and the order in which they appear (top to bottom) is randomized. This is an example of a forced-choice test from psychophysics. Once a button is clicked the next sentence is played, thus starting the next trial. It is not possible to replay any sentence.

If the spoken sentence ends with the word "bed" and the recording has high intelligibility, the word "bed" is easy to distinguish and the vast majority of the trials will lead to the selection of the correct answer. If the recording has low intelligibility, subjects may hear "wed," "fed," "shed," "led," or "red" instead of "bed" and many of the trials will lead to the selection of an incorrect answer. MRT trials are performed repeatedly on each condition and each trial is classified as a success or a failure. The success rate is corrected for the rate at which correct words could be guessed (1/6) and this then becomes the MRT intelligibility value. MRT values range from 0.0 to 1.0, or alternately 0% to 100%. In either case, the low end of the scale indicates a success rate that is at the guessing level, and the high end of the scale indicates perfect success at identifying words.

3.4.2   ABC-MRT

MRT design, implementation, and analysis can require considerable time and effort. Systems are often tested using hundreds or thousands of trials and dozens of listeners. This can require weeks or months of effort as well as a suitable listening environment. This

motivated the development of a signal processing based alternative, Articulation Band Correlation Modified Rhyme Test (ABC-MRT)[8][9]. This uses the same speech material as the true MRT input to the SUT, but also uses an objective algorithm to analyze the output of the SUT. This eliminates the need for human listeners and a controlled listening environment and allows testing to happen in real time. The savings are dramatic, but of course ABC-MRT can only provide estimates of true MRT values. An update to ABC-MRT, called ABC-MRT16, was published in 2017, which upgrades the algorithm to be more flexible and work with higher bandwidths. In this paper ABC-MRT16 was used for all testing, and all mentions of ABC-MRT from this point refer to ABC-MRT16.

ABC-MRT uses MRT speech recordings and performs a speech recognition task that is analogous to the task in the actual MRT. Unlike most speech recognition algorithms, the ABC-MRT algorithm does not strive for maximal robustness to noise and distortion. Instead it strives for a robustness that is similar to that of human listeners. ABC-MRT uses temporal correlations within articulation index bands to select one of six possible words from a list. The rate of successful word identification becomes the measure of speech intelligibility, just as in the MRT. Because the robustness is similar to that of humans, the ABC-MRT scores are similar to those of humans. Thus, it can be said that ABC-MRT provides estimates of true MRT values.

For each trial (e.g., "Please select the word 'bed'") the ABC-MRT algorithm produces a single value, nominally in the range from zero to one. This value is analogous to those produced by the MRT. An output value of zero corresponds to no conveyance of speech information. Any values below zero are equivalent to zero and are indicative of estimation error. An output value of one corresponds to perfect conveyance of speech information and indicates that the correct word is identified on every trial. The correlation between ABC-MRT16 results and actual MRT results across 367 SUTs was measured at 0.95 in Ref. [9]. In this report ABC-MRT is applied in some novel situations that are outside the range covered in Ref. [9]. It is possible that previous intelligibility estimation performance might not be realized in the current work, thus follow-on work is planned to address any possible limitations due to this fact.

The MRT testing paradigm must be modified before it can be leveraged in access delay measurements. MRT results are driven by the keyword that follows the carrier phrase and thus is the *last* word in a message. However, access delay measurements are inherently driven by the *first* word in a transmitted message. Thus, it was necessary to move to a modified MRT testing paradigm that uses only keywords and no carrier phrases. The modified ABC-MRT is part of the code package available at https://doi.org/10.18434/M32085.

4. Technical Approach

The access delay measurement system presented here is a technology-independent measurement system. The system is designed around the basis of user experience in voice communications: the speech being transmitted and received. As such, it allows for fair comparisons of performance of any speech-based communications technology. The devel-

oped system is a single location measurement system, where both the transmit and receive communications devices are in the same location. The system relies on speech intelligibility measurements to determine access delay. Thus, it is ideal to perform measurements in a controlled RF environment to minimize the impact of the signal quality on intelligibility scores. More specifically, variations in the RF environment will cause variations in speech intelligibility and these variations will confound the measurement processes described below.

4.1 Defining the Problem

Here end-to-end access time is defined as follows:

Definition 1 : ***End-to-end access time*** *is the total amount of time from when a transmitting user first presses PTT until a receiving user hears intelligible audio.*

This definition is composed of two components: M2E latency and access delay. This definition is different from how access time is conventionally discussed in 3GPP and TIA standards, as it is driven by the actions (button presses) and experiences (hearing intelligible audio) of the users.

Access delay is defined as the minimum length of time a user must wait between pressing a PTT button and starting to speak to ensure that the start of the message is not lost. This definition is immediately more ambiguous and more difficult to define than that of M2E latency. It is challenging to characterize whether a message is lost or not. To address this issue, ABC-MRT is used to measure the intelligibility of audio in the measurement system. The definition of access delay is formalized more rigorously as:

Definition 2 : *Access delay is the minimum length of time a user must wait between pressing a PTT button and starting to speak to ensure that the first word of the message has an average intelligibility that is no lower than $\alpha \cdot I_0$, where $0 < \alpha \leq 1$, and I_0 is the baseline intelligibility of that word through the communications system.*

This definition of access delay is different from other access time definitions that focus primarily on the time for the system to allocate channels for users. Instead, this definition encompasses all the aspects in a communications technology that contribute to the delays between two users hearing each other. This includes the channel allocation time, but also time for components on both the transmitting and receiving devices to turn on and to be fully operational for sending and receiving voice. The different components that contribute to M2E latency and access delay are shown in Table 1.

It is worth noting that in actual use, access delay time will include a user-induced pause. Depending on the system, this is either the pause between pushing the PTT button and

Table 1. End-to-end Access Components

Term	Transmit Device	System	Receive Device
M2E Latency	Front End Audio Processing Speech Coding Communications Stack (i.e., Data Interleaving, Forward Error Correction)	Signal Propagation Routing Control (i.e., Channel Assignment, Direction) Backhaul (Fiber/Microwave) Transcoding	Communications Stack (i.e., Data De-interleaving, Forward Error Correction) Speech Decoding Audio Processing
Access Delay	Transmitter Start Up Access Request Process Grant Speech Processing Start Up	Signal Propagation Access Request Processing and Granting (i.e., Database Credentialing, Channel Availability, Locate Talk Group Users) Signal Propagation	Channel Acknowledgment

speaking, or the pause between hearing the "access granted beep" and speaking. For efficient communications, the user will strive to make this pause as short as possible, but will need to also include some safety margin to avoid miscommunications or retransmissions. The access delay measurement system presented here is designed to give that shortest possible value, subject to a certain level of average intelligibility.

This definition of access delay is also applicable to direct mode communications as well as infrastructure modes. Direct mode communications are generally outside the scope of other access time definitions and measurements.

4.2 Measurement Concept

The measurement system is designed based on definition 2. In particular, it focuses on repeatedly sending a predefined audio clip through a communications system while varying where in the clip PTT is triggered, referred to as PTT time. By measuring the relationship between PTT time and the intelligibility of the first word of the audio clip, access delay can be determined.

4.2.1 Audio Clips

Audio clips are designed as follows: a single word from the database of ABC-MRT words [9] is selected. Words are selected only from the MRT Audio Library [10] batches where the leading consonant varies. For example, the word *went* might be selected from the batch of words {*went, sent, bent, dent, tent, rent*}. Missing the beginning of these words makes it difficult to discern which word from the batch was spoken. This places the majority of the emphasis from an intelligibility standpoint on the beginning of the word. In general, it is unlikely that the most vital portion of a message is the very first part of the first word, so the resulting measurements are worst-case.

The clip is structured with T seconds of silence in the beginning, where T is chosen

11

so that system access time is almost assuredly less than T seconds[2]. The word is then played, followed by T seconds of speech and/or silence; this is called the filled section and is described below. After the filled section, the word is played again. The first play of the word is called P_1 and the second P_2. An example audio clip with T set to two seconds can be seen in Fig. 5.
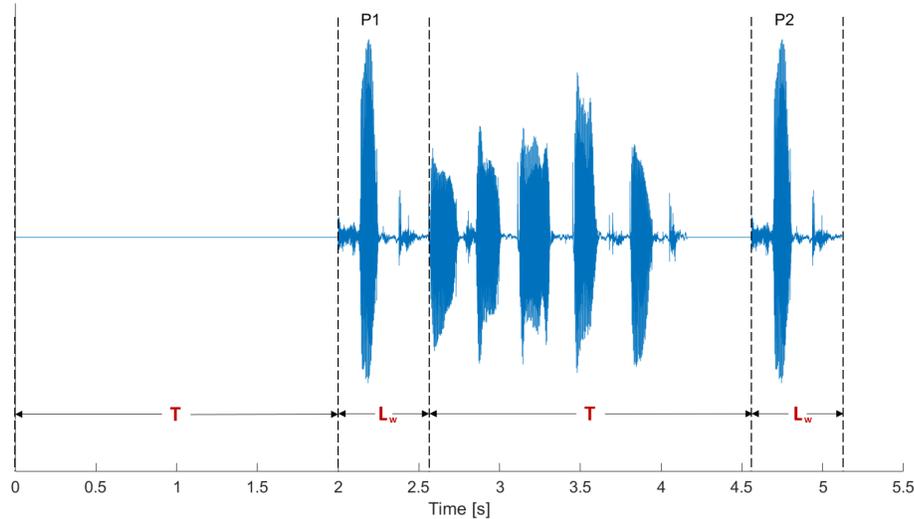
Fig. 5. Example transmit audio with $T = 2$ seconds.

The speech in the audio clip between P_1 and P_2 is referred to as the filled section of the clip. It is generated via MRT words that come from batches where the ending consonant varies, to avoid having clips from the same batch as that of the words under test. The filled section is composed of the shortest words that meet the previous criteria and are of a total length less than $T - 0.1$ seconds. The 0.1 seconds serves as padding to preserve the intelligibility of P_2. The filled section is always T seconds, and thus can contain silence at its end, right before P_2.

The filled section of the clip was incorporated for two reasons. Firstly, it more accurately mimics real world voice communications; typical conversations do not contain a single word followed by a long period of silence before the next. Secondly, audio clips that are more dense with speech allow for more accurate calculations of M2E latency.

4.2.2 PTT Timing

The test then involves repeatedly sending the clip through a communications system while varying when PTT is pressed relative to the start of the clip. PTT times can be varied from

---

[2] In initial testing, certain technologies occasionally exhibited much longer access times than average. It is thus recommended to select T to be up to twice as large as access time is expected to be. If there is no prior knowledge of system access time, T should be chosen conservatively.

the beginning of the clip up to the end of P_1. In other words, if the length of the word is defined by L_w, then PTT times can be varied from 0 seconds until $T + L_w$ seconds. The amount of time between PTT times are referred to as time steps. Figure 6 shows examples of received audio for three different PTT times. When the PTT time is large the system does not have enough time to grant access, and the speech of P_1 is either lost completely, or partially cut off. When the system is given enough time to set up completely, P_1 is sent in its entirety with no issue.

Each time the clip is sent through the communications system, the intelligibilities of P_1 and P_2 are calculated. Due to the way that T is selected, the intelligibility of P_2 represents the asymptotic intelligibility, I_0, of the word. During a measurement, PTT times start at $T + L_w$ seconds and increment towards smaller values. When the intelligibility of P_1 reaches that of P_2 the test is stopped.
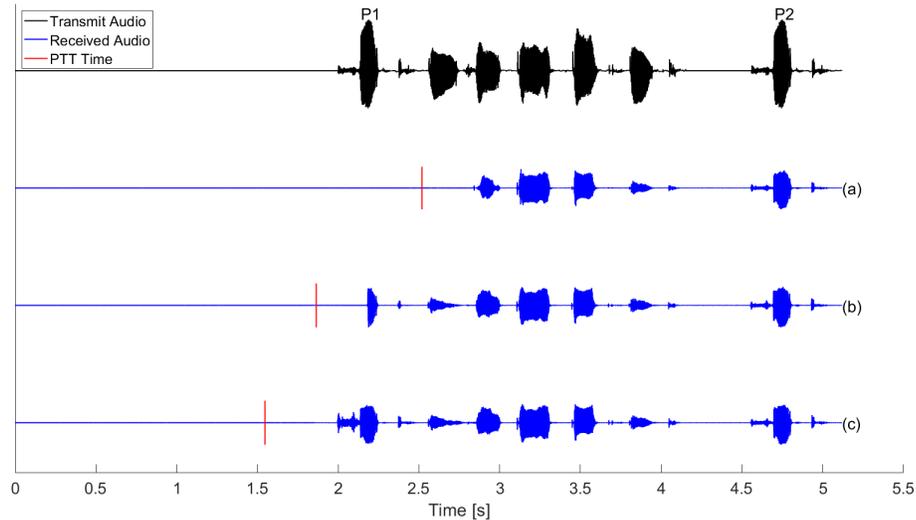
Fig. 6. The black waveform shows the full transmitted audio clip. The red vertical lines show where in time PTT was pressed. (a) PTT was pressed after P_1 and no audio is received for P_1. (b) PTT was pressed shortly before P_1 and the system was able to partially set up in time to transmit some but not all of P_1. (c) PTT was pressed a safe time before P_1 and the system was able to fully set up in time to transmit all of P_1.

For a measurement, multiple trials must be performed at each PTT time. This helps to characterize the variability in system access delay and the variability in intelligibility. More repetitions performed at a given PTT time yields more informative average intelligibilities. This comes at the cost of time, as it generally takes at least five seconds to perform a single trial. Furthermore, the resolution of the PTT times is key to developing a meaningful notion of access. If PTT times were stepped in half second increments, there would be less detailed information on when access was granted. The time step defines the resolution of the measurement.

13

Both the resolution of the measurement and the accuracy at each PTT time must be balanced with the time it takes to perform trials to achieve useful access delay measurement results.

4.2.3 Intelligibility Curves

Before fitting the data to a curve, PTT times must be transformed in a way that is not dependent on the structure of a given audio clip and that relates directly to definition 2. In this way data from multiple clips can be combined. As discussed in Sec. 4.2.1, audio clips are designed with T seconds of silence before P_1 and T seconds of filled section before P_2. In order to relate to definition 2, a time axis that shows the time between when PTT was pressed and when voice first arrived at the transmit device is used. For example, when PTT is triggered at the beginning of P_1 (T seconds into the audio clip) this becomes 0 seconds in the new time scale. Any PTT times that happened prior to the start of the word (less than T seconds) are positive times in the new time scale. Conversely, PTT times that happened after the start of the word (greater than T seconds) are negative times. If the PTT time in audio clip time is represented by $T_{ptt} \in [0, T + L_w]$ for an audio clip with T seconds of silence and/or speech preceding P_1 and P_2, the new time scale representation, t, is given by

$$t = T - T_{ptt} \, , \tag{1}$$

with $t \in [-L_w, T]$.

The new time scale, t, relates directly to the user experience. It corresponds to the waiting time between pressing PTT and starting to speak. Larger values of t give the system more time to set up and thus improve the chances of intelligible speech at the receive side. For most efficient communications, users will naturally seek the smallest value of t that reliably produces intelligible speech at the receive side. A t value of zero corresponds to no waiting and negative values correspond to speaking before pressing PTT. Figure 7 illustrates three examples of T_{ptt} and how those values are reflected in t.

Once the intelligibility of P_1 has been measured across different PTT times, a curve can be fit that represents intelligibility as a function of PTT time relative to word start. There are a large variety of curves that could model the transition from unintelligible to intelligible based on PTT time. Here the data is fit to a logistic curve. An example logistic intelligibility curve is shown in Fig. 7.

Once intelligibility has been modeled as a function of time, $I(t)$, access delay can be calculated for any value of α. Note that regardless of how much time elapses between PTT being pressed and speech being sent to the transmit device, intelligibility on average will be bounded between 0 and the asymptotic intelligibility of the system, I_0. $I(t)$ represents the average intelligibility of the system when t seconds are given between PTT being pressed and speech being sent to the transmit device. Given $0 < \alpha < 1$, an intelligibility of $\alpha \cdot I_0$ can be achieved with $t = I^{-1}(\alpha \cdot I_0)$. Access delay can then be defined as

$$\tau_A(\alpha) = I^{-1}(\alpha \cdot I_0) \, . \tag{2}$$
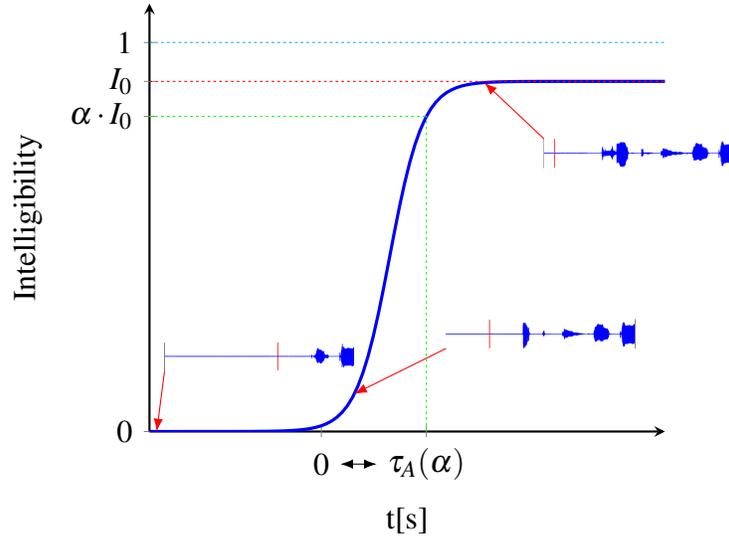
Fig. 7. Here I_0 represents the asymptotic intelligibility. Access time values are defined as a function of α, where $0 < \alpha < 1$ describes the fraction of asymptotic intelligibility at which access is granted. Waveforms with varying T_{ptt}, as in Fig. 6, are shown with their corresponding locations on the intelligibility curve.

When fitting to a logistic curve the intelligibility function takes the form:

$$I(t) = \frac{I_0}{1 + e^{(t-t_0)/\lambda}} \ . \tag{3}$$

This then yields the access delay function

$$\tau_A(\alpha) = \lambda \ln\left(\frac{1 - \alpha}{\alpha}\right) + t_0 \ . \tag{4}$$

Thus the access delay function is parameterized by just two terms output from fitting the logistic curve: λ and t_0. The term λ determines how abruptly the intelligibility transitions from 0 to I_0 and the term t_0 gives the midpoint of that transition ($I(t_0) = I_0/2$).

4.3 Measurement System Design and Setup

The test setup requires one laptop, one microcontroller, and one audio interface, as well as the two communication devices on which the access time measurement is being performed. The computer sends audio out via USB to the audio interface which then outputs an analog audio signal, through a playback jack, to the transmit device. The receive device outputs the received analog audio signal to the audio interface. The audio interface digitizes the signal and passes it to the laptop where it is then recorded in MATLAB.

The measurement system relies on using MATLAB to both play and record audio out of an audio interface in order to simultaneously send audio to be transmitted while recording

15

received audio. It is important to note that there is a non-negligible amount of time required for the signal to travel through the measurement system itself, even when no SUT is present.

Figure 8 shows a diagram of the test setup. The measurement system sends audio and the PTT signal to a transmit device while recording received audio and the PTT confirmation tone. A microcontroller is attached to the computer via USB and to the transmit device in order to place the device into transmit mode at the correct time. To accomplish this the microcontroller uses a time-reference derived from a sine wave "start tone" (from the audio interface) along with a time-offset value that is provided by MATLAB via the USB connection.
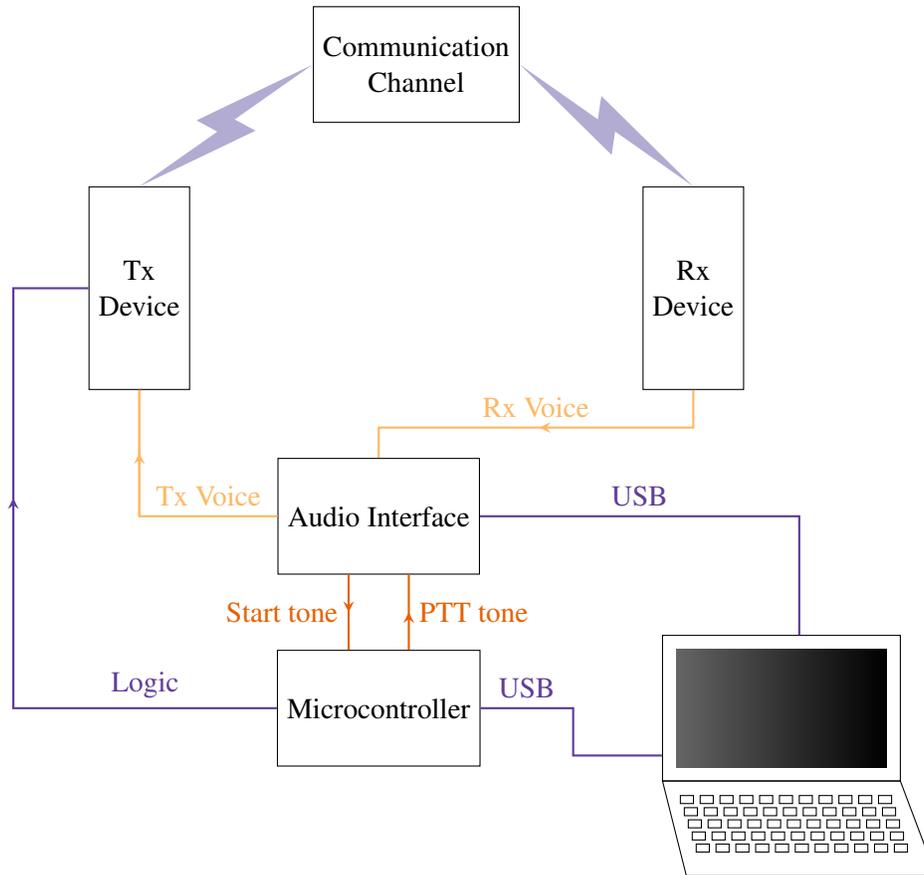
Fig. 8. Diagram of measurement system setup

For the most reliable and meaningful measurements, the audio interface must be high-quality and provide continuous streams of samples without missing any samples or inserting any erroneous samples. Audio interfaces with adjustable buffer sizes and USB streaming settings offer more control to achieve this goal. The audio interface must be able to simultaneously play and record audio via the MATLAB function *audioPlayerRecorder* in the Audio System Toolbox(2018)/Audio Toolbox[3]. In addition to the Audio Toolbox, this

---

[3]Mathworks 2019, https://www.mathworks.com/help/audio/index.html

measurement system requires the Signal Processing Toolbox[4].

Many USB audio interfaces meet these requirements; all tests discussed in this paper used Behringer UMC404HD and UMC204HD audio interfaces with Behringer driver version 3.29.0. The buffer size was set to *512* samples and the USB Streaming set to *Standard* to avoid buffer over/under runs and audio glitches, respectively. This setting will vary by setup and therefore ideal settings will be system dependent.

If one is utilizing the setup from the M2E latency measurement system [1], additional information on updating that setup to complete access delay testing can be found in Appendix B.

4.3.1 Microcontroller

For access time measurements, keying the radio is mandatory. A microcontroller is attached to the computer via USB and to the transmit device. An MSP430F5529 Launchpad was used to trigger PTT with a few additions regarding timing in order to place the device into transmit mode at the correct time. The MSP430 is a microcontroller family sold by Texas Instruments. The launchpad board used has a MSP430F5529 microcontroller as well as an in-circuit debugger and USB hub. The MSP430F5529 runs at up to 25 MHz, has an integrated USB interface as well as 8 kB on chip RAM and 128 kB on chip Flash.
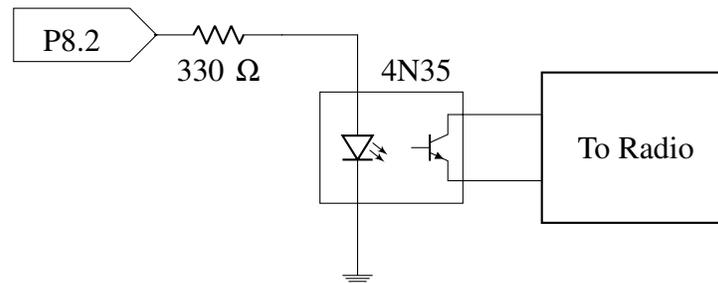
Fig. 9. Circuit schematic for optoisolator used for PTT triggering

An optoisolator circuit was used to interface to the radio as shown in Fig. 9. To turn the PTT on, the microcontroller brings P8.2 (port 8, pin 2) high, which turns on an optoisolator connected to the PTT wires of a headset connected to the radio. Figure 10 shows the microcontroller with the optoisolator circuit attached. The connections to power and I/O are accomplished through two dual row headers on the board. To control the optoisolator, P8.2 was used. The components were soldered to a perfboard to allow easy modifications.

Originally, the microcontroller activated PTT at a specific time after receiving a command. This approach produced inconsistent results because it was not possible to maintain a constant time relationship between sending a command to the microcontroller and starting audio playback. This problem was solved by an audio time reference in the form of a sine wave. The MATLAB code sends out a sine wave (shown as "start tone" in Fig. 8)

---

[4]Mathworks 2019, https://www.mathworks.com/help/signal/index.html

17
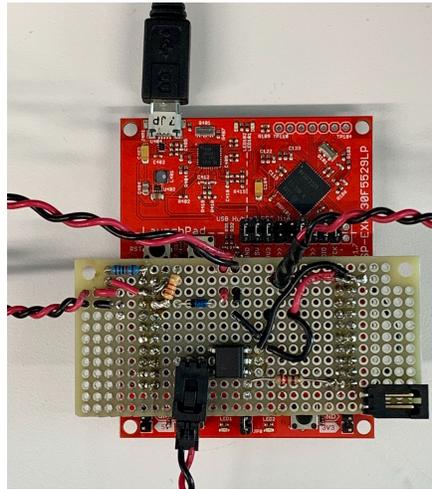
Fig. 10. Microcontroller with added hardware for use with access time

on channel 2 of the audio interface at the start of the clip. Because this tone appears at the beginning of the clip, it occurs T seconds before any audio is presented to the transmitting device. This sine wave signals the microcontroller that the audio playback has started. Figure 11 shows the circuit used to allow the microcontroller to detect this sine wave start signal.
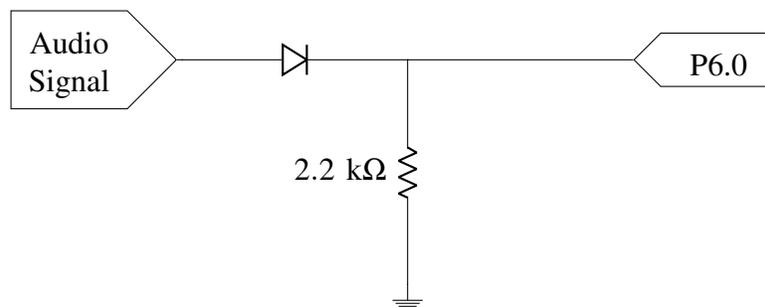
Fig. 11. Circuit schematic for start detector for accurate PTT timing. This circuit uses a rectifier to detect the envelope of the signal and starts timing after the envelope exceeds a threshold.

The circuit rectifies the incoming sine wave and uses the internal comparator of the microcontroller to detect the rising edge of the audio signal. The comparator threshold is set in the microcontroller so that when the start signal goes over 0.625 V, the start signal is received. A timer is used in capture mode to record when the edge occurs and then used in compare mode to trigger an interrupt when the PTT needs to be pushed. The time offset between when the start tone is detected and when PTT is to be pushed is sent from MATLAB to the microcontroller in advance of this process. The timing from this setup was found to be much more consistent and allowed the PTT time to be set closer to the beginning of the clip.

The system also provides feedback to indicate when PTT was actually pressed. The microcontroller outputs a square wave signal, on P2.5, at the same time that it activates PTT. This square wave is captured by a second input channel on the audio interface and thus provides a record of the exact actual PTT timing. This signal is shown as "PTT tone" in Fig. 8.

The characterization of the measurement system delay is discussed in Sec. 5.2, and allows for measurements to be corrected to reflect only the delay of the communication devices. The measurement system latency must be known in order to correct the measured PTT start times.

4.4 Maximizing Speech Quality

Controlling speech quality is a key component to designing a technology-independent QoE measurement system. Different technologies may have varying levels of processing capabilities to enhance the quality of speech transmitted, depending on what they receive. Rather than trying to disable these features in technology before testing them, it is best to test and compare them in the operating status in which they will be commonly used. However, if the quality of the speech being sent is not accounted for, tests may not be repeatable as devices transition into different operating modes. Therefore, a fair test must send speech of the same quality and format to every technology tested, so that they are compared under the same circumstances.

4.4.1 Ground Loop Effects

The access delay test setup requires the use of two transformers for cabled testing in order to reduce noise caused by a ground loop. A 1:1 transformer is located between the receive radio output and the audio interface input. The transmit radio path requires the use of one 12:1 direct box transformer located between the audio interface output and transmit radio input. More information on the effects of a ground loop on the test setup can be found in Ref. [1].

4.4.2 Audio Volume

Two of the primary factors that affect the quality of speech are the volumes at which speech is transmitted and received. Transmit volume, denoted V_{Tx}, is controlled via the audio interface and defines the power of the signal sent from the audio interface to the transmit device. It mimics the volume of a user speaking into a microphone. If V_{Tx} is set too loud, like when a user screams, audio would over-drive the transmit device and quality would go down. Similarly if it is too quiet, like when a user whispers, the signal would get lost in the noise and be indiscernible.

Receive volume, denoted V_{Rx}, can be controlled in two areas: on the receive device and the input gain knob on the audio interface. It is recommended to fix the volume setting on a receive device to a single setting and only adjust V_{Rx} via the input gain knob on the audio

interface. Adjusting just one of the V_{Rx} control knobs allows for consistency in variable settings. When V_{Rx} is set too high, the recorded audio will be clipped, severely impacting quality. If it is too low, the recorded signal will be difficult to distinguish from system noise.

Instructions for the volume setting procedure can be found in Appendix A.2. V_{Tx} and V_{Rx} are susceptible to being interdependent, so the procedure is an iterative approach. Audio quality can be determined using audio quality measurements such as PESQ [11]. Note that the procedure closely resembles that of Ref. [1], but has been adapted to focus less on P25 technologies and to be flexible for any communication system.

4.5 Uncertainty Calculations

In order to calculate the uncertainty of the access delay times returned by the measurement system, Eq. (4) is used. In particular, an access delay estimate, $\hat{t}$, for some choice of α can be defined as

$$\hat{t} = \hat{\lambda} \ln\left(\frac{1-\alpha}{\alpha}\right) + \hat{t}_0 \ . \tag{5}$$

Let $C = \ln\left(\frac{1-\alpha}{\alpha}\right)$. It can then be seen that

$$\mathrm{Var}(\hat{t}) = C^2 \mathrm{Var}(\hat{\lambda}) + \mathrm{Var}(\hat{t}_0) + 2C \cdot \mathrm{Cov}(\hat{\lambda}, \hat{t}_0) \ . \tag{6}$$

The uncertainty of $\hat{t}$, $u(\hat{t})$, is then $u(\hat{t}) = \sqrt{Var(\hat{t})}$. Thus, the uncertainty of the access delay estimate can be determined from just α and the variance and covariance of the estimates returned by the logistic curve fit. The estimates, as well as their variance and covariance, are calculated using the *minpack.lm* package in *R* [12]. Equation (6) is utilized when calculating the uncertainty of access delay measurements in Sec. 5.2 and Sec. 5.3.

5. Example Access Delay Measurements

Example access delay measurements were performed on digital P25 and analog LMR systems. The P25 system was tested in direct as well as Phase 1 and Phase 2 trunked modes, while the analog system was tested in direct and conventional modes. All tests were performed in a controlled lab setting. Depending on the communications devices utilized and their power ratings, it was sometimes necessary to add attenuators in line with RF cables. For consistency, all tests were performed on the same model of radio operating in the appropriate communication mode.

5.1 Additional Measurement Information

In order to perform tests on the LMR systems, external 3-wire microphone/earphone/PTT devices were modified, as seen in Fig. 12. These connectors form the connections, shown in Fig. 8, between the audio interface and the radios as well as between the Tx radio and

the microcontroller. These adaptations allow PTT to be triggered via microcontroller commands, speech to be sent from the audio interface output to the transmit radio input, and speech to be sent from the receive radio output to an audio interface input.


(a) Unmodified headset:
(1) PTT
(2) Microphone
(3) Ear-piece

(b) Headset modified for transmit:
(1) PTT connection
(2) RCA connection for transmit audio
(3) Ear-piece

(c) Headset modified for receive:
(1) PTT
(2) Microphone
(3) XLR connection for received audio

Fig. 12. Adapted LMR headsets

A trial consists of a single instance of playing an audio clip, pushing PTT, and recording the received audio. A session consists of some number of consecutive trials. Sessions can use a single audio clip or alternate between different audio clips. The number of trials in a session is dependent on the number of audio clips used, the size of the time step used in the measurement, the number of repetitions per PTT time, and the SUT.

Speech intelligibility is always a function of an individual's voice characteristics as well as the SUT. Thus, testing with several different voices is more robust than testing with just one voice. Access delay tests were performed with four speakers, two male and two female, each speaking a different word. The speakers were labeled *F1, F3, M3,* and *M4* and their respective words were *hook, west, cop,* and *pay*. These words were chosen to give the smallest root mean square error (RMSE) between the average ABC-MRT scores for these four words and the corresponding ABC-MRT score when all 1,200 MRT words are used. The RMSE operation was performed across 139 radio and codec conditions described in Ref. [8]. The RMSE for the four selected words is 0.1097, and corresponds to the smallest RMSE possible under the conditions that one word is selected from each of four speakers, and words are only drawn from batches with differing leading consonants.

For all the tests below, time steps were set to 20 ms. At each PTT time, 30 trials were repeated for each audio clip. These selections were chosen to balance accuracy of the measurements with the high time cost associated with performing the measurements.

To further limit the time cost of measurement, an automatic stopping condition was enabled. Tests were started with PTT times at the end of P_1 and then PTT time decreased until the intelligibility of P_1 matched that of P_2. The intelligibility of P_1 was compared

to that of P_2 via the approximate permutation test with 10,000 resamples [13]. The intelligibility values for all the repetitions of P_1 at a given PTT time were compared to all measured intelligibilities of P_2 across the whole test. If the p-value from the approximate permutation test for a given PTT time was less than 0.05, a flag was thrown. To limit the stop condition's susceptibility to false alarms, the test did not stop until flags were thrown at six consecutive PTT times. The number of consecutive PTT times required for the test to stop is an adjustable parameter. The only negative side effect of setting this parameter too high is the time cost of the measurement, so it is recommended to use a conservative value to ensure the test does not stop too early. Furthermore, the value of this parameter should be made larger when smaller time step values are used to protect against the potential for ABC-MRT mischaracterizations when passing certain fractions of a word.

Data was collected with separate sessions for each audio clip. The number of trials in a session could be huge, so the test was set to pause every 300 trials. The pause provided a chance to check the test status, replace radio batteries if needed, and save test data in case of extended breaks.

It is worth noting that all measurements were performed on single site systems with low numbers of users. Access delay is likely related to the number of users on the system but has not been proven in this measurement effort because it is beyond the current scope of this project.

5.2 Measurement Characterization

The measurement system M2E latency must be characterized in order to understand the inherent latency of speech through the system. This allows for the access delay PTT times and the M2E latency values to be corrected to only represent the performance of the devices themselves.

It is also important to characterize the measurement system's response to the simplest possible voice communications system. This is a system that does not pass speech until the PTT signal is received, with minimal turn-on time. Such a system provides lower bounds on possible access delay measurements from the measurement system. The PTT gate was designed to meet these criteria; integration of this device within the test system can be seen in Fig. 13. The results of this characterization describe the effect that muting portions of words has on speech intelligibility. Measurements using the PTT gate yield the necessary M2E latency and access delay characterizations.

5.2.1 A Simple Replacement for a Radio System

Characterizing the access delay measurement system required the addition of a device, referred to as the PTT gate, that granted access in a consistent manner. This device, shown in Fig. 14, consisted of an AQV210EH PhotoMOS electronic switch to block the audio until the PTT was pushed. To prevent crosstalk, the audio output of the switch was terminated using a MOSFET switched resistor to ground when the PTT was not pushed.

22
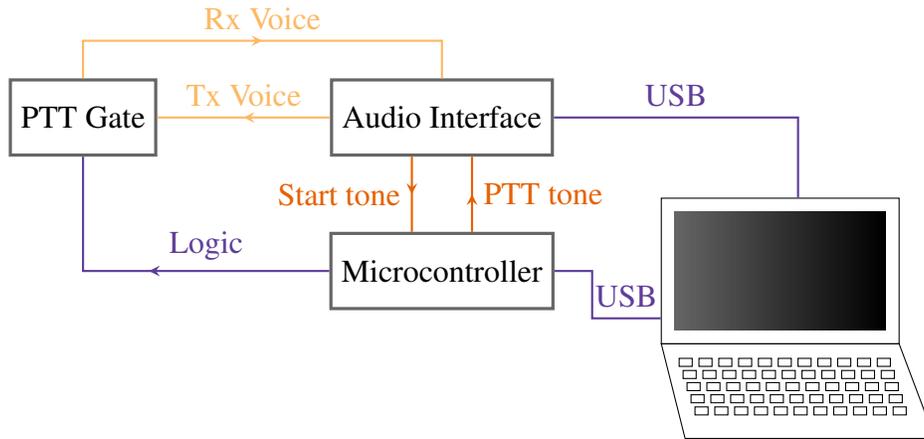
Fig. 13. PTT gate test setup

The AQV210EH has a maximum turn-on time of 2 ms. However, in lab testing the average observed value and 95% confidence interval was $93.708 \pm 0.005\ \mu$s. As the device consists of a LED and a photo transistor inside an IC, the turn-on time is expected to be consistent run-to-run.
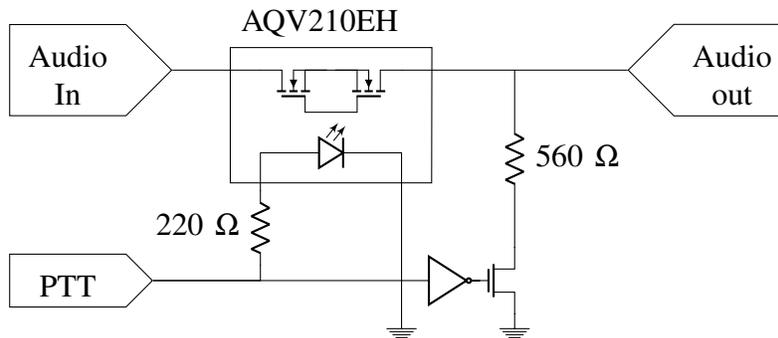
Fig. 14. Circuit schematic PTT gate

5.3 Measurement Results

Table 2 details the parameters of the intelligibility and access delay curves generated from the example measurements. The three parameters, I_0, t_0, and λ, fully describe the intelligibility logistic curves detailed in Eq. (3). Many features of the tested technologies can be immediately understood from Table 2 and can be visually confirmed in Fig. 15. Only t_0 and λ are needed to define access delay as seen in Eq. (4). This section primarily discusses these measured parameters and presents some example access delay results based on an intelligibility that is 90% of the asymptotic intelligibility level ($\alpha = 0.9$). This selected intelligibility level is arbitrary and results for more asymptotic levels as well as access delay comparisons at a fixed intelligibility level are included in Appendix D.

23

Uncertainty values for all access delay measurements were calculated using Eq. (6). All reported uncertainty values represent a 95% confidence interval with a coverage factor of $k = 1.96$. It is worth noting that the magnitude of the measurement uncertainty compared to the magnitude of the measurement results is relatively small, as can be seen clearly in Fig. 16. This shows that the measurement system is well-suited to comparing different technologies, as the results are reported with enough confidence to make distinctions at a resolution that is likely lower than what the human ear would be able to perceive.

Table 2. Curve Parameter Results with 95% confidence intervals included in parentheses

	I_0	t_0 [s]	λ [s^{-1}]
PTT Gate	$1.0, (1.0, 1.0)$	$-0.120, (-0.122, -0.118)$	$-0.028, (-0.029, -0.026)$
Analog Direct	$0.984, (0.983, 0.984)$	$0.071, (0.069, 0.073)$	$-0.035, (-0.037, -0.034)$
Analog Conventional	$0.984, (0.984, 0.984)$	$0.224, (0.223, 0.226)$	$-0.033, (-0.035, -0.032)$
P25 Direct	$0.965, (0.963, 0.966)$	$-0.006, (-0.008, -0.004)$	$-0.039, (-0.041, -0.037)$
P25 Trunked Phase 1	$0.973, (0.971, 0.974)$	$0.534, (0.531, 0.536)$	$-0.055, (-0.058, -0.052)$
P25 Trunked Phase 2	$0.910, (0.905, 0.915)$	$0.521, (0.519, 0.524)$	$-0.064, (-0.067, -0.062)$

The parameter t_0 describes the 50% of asymptotic intelligibility point for each technology. In other words, t_0 describes the amount of time on average a user must wait between pressing PTT and speaking for 50% of the asymptotic intelligibility of the technology to be achieved on the receiving end. Thus, t_0 provides a general sense of access delays for each technology. Access delay is further described by λ, which describes the steepness of the intelligibility transition from unintelligible to the asymptotic level. In particular, the smaller the value of λ, the slower the transition from unintelligible speech to the asymptotic level. In this application, λ is always negative, so smaller values are values with greater magnitude. The parameter λ captures both the variability in intelligibility for a technology as well as the variability in access delay. Larger degrees of variability in each of these components will lead to smaller values of λ. It is natural to assume that technologies with more infrastructure and actual channel assignment processes will have more variability in their setup time. Thus, it is expected to see λ be smaller for such technologies.

These parameters and their implications can first be explored by considering the PTT gate. The PTT gate is a fast switch with virtually no distortion or variation. Since the PTT gate has this known behavior, measurements of the PTT gate provide insight into the measurement system itself, and are a good starting point for understanding the behavior of the measurement system parameters. In fact, the PTT gate results provide upper bounds for I_0 and λ and a lower bound for t_0. The minimal distortion in the PTT gate results in intelligibility values of 1. Similarly, the lack of technology in the PTT gate results in the steepest possible transition from unintelligible speech to intelligible speech, leading to an upper bound on the steepness parameter, λ. The lack of technology and variation in the PTT gate also results in a lower bound on t_0, as the PTT gate represents the simplest possible PTT communication system. As can be seen in Fig. 15, the entire transition from unintelligible to intelligible voice occurs prior to the start of the word, $t = 0$ s. The negative access delay values for the PTT gate, seen in Fig. 16, demonstrate that it is not necessary

24

to receive all of a word in order to understand it. The 90% asymptotic intelligibility access delay result is $\tau_A(0.9) = -59.4 \pm 3.9$ ms. This has the interpretation that on average one can miss up to the first 59 ms of the MRT keywords used here and still achieve an intelligibility of at least 90%.

Both analog direct and conventional modes lack the infrastructure generally associated with access time measurements. There is no channel allocation as there is only a single frequency required in direct mode and there are dedicated transmit and receive frequencies in conventional mode. Both technologies had high asymptotic intelligibility levels, each achieving values of 0.984. Similarly, both had relatively high λ values, that were close to that of the PTT gate. This is due to the lack of variability in the setup times of these technologies; analog direct has no infrastructure while the infrastructure for analog conventional involves no channel assignments. The values of t_0 differed quite a bit however, with analog conventional having a value of 224 ms, about triple that of analog direct with its value of 71 ms. This can be explained by the additional infrastructure in the conventional setup. Once the transmit device sends the message along in direct mode, the message goes straight to the receiving user. In conventional mode the transmit device sends the message to a repeater, which must receive and re-transmit the message. It is likely that the repeater is not immediately prepared to perform this task, and the increase in t_0 between the technologies is caused by the setup time of the repeater. Thus, a user of the analog conventional system tested would need to account for the setup time of the repeater and wait an additional amount of time between pressing PTT and speaking to ensure that the message is transmitted intelligibly. The 90% asymptotic access delay result for analog direct was $\tau_A(0.9) = 148.8 \pm 3.8$ ms and for analog conventional it was $\tau_A(0.9) = 297.8 \pm 2.8$ ms.

Similarly to analog direct mode, P25 direct mode has no communications infrastructure. This contributes to its very low value of t_0. The value of λ is slightly lower for P25 direct mode than either analog mode. This could be attributed to the use of a vocoder to digitize the speech for transmission. Over the course of testing, it was observed that the measurement system would record received speech that was highly unintelligible. Furthermore, it was difficult to reproduce the scenarios that caused this garbled speech. It was hypothesized that this was caused by the way vocoder frames aligned with portions of the partially muted words, causing the vocoder to occasionally send through speech that did not resemble the transmitted audio. The smaller λ value supports this idea, as the susceptibility of the vocoder frames aligning with speech introduces additional variability to how speech is transmitted through the technology. The impacts of the P25 vocoder can also be seen in the intelligibility of the system, with its lower asymptotic level of 0.965, and the larger spread of data points at the asymptotic level in Fig. 15. The 90% asymptotic access delay result for P25 direct mode was $\tau_A(0.9) = 79.1 \pm 4.4$ ms.

The measurement system was also tested on P25 trunked Phase 1 and Phase 2 systems. Both of these communication modes have the channel requests and grants generally associated with access time measurements. Thus, it was immediately expected that they would have much larger access delays than any of the other example measurements performed. The impacts of the channel grant process is seen in both t_0 and λ. The fact that an access
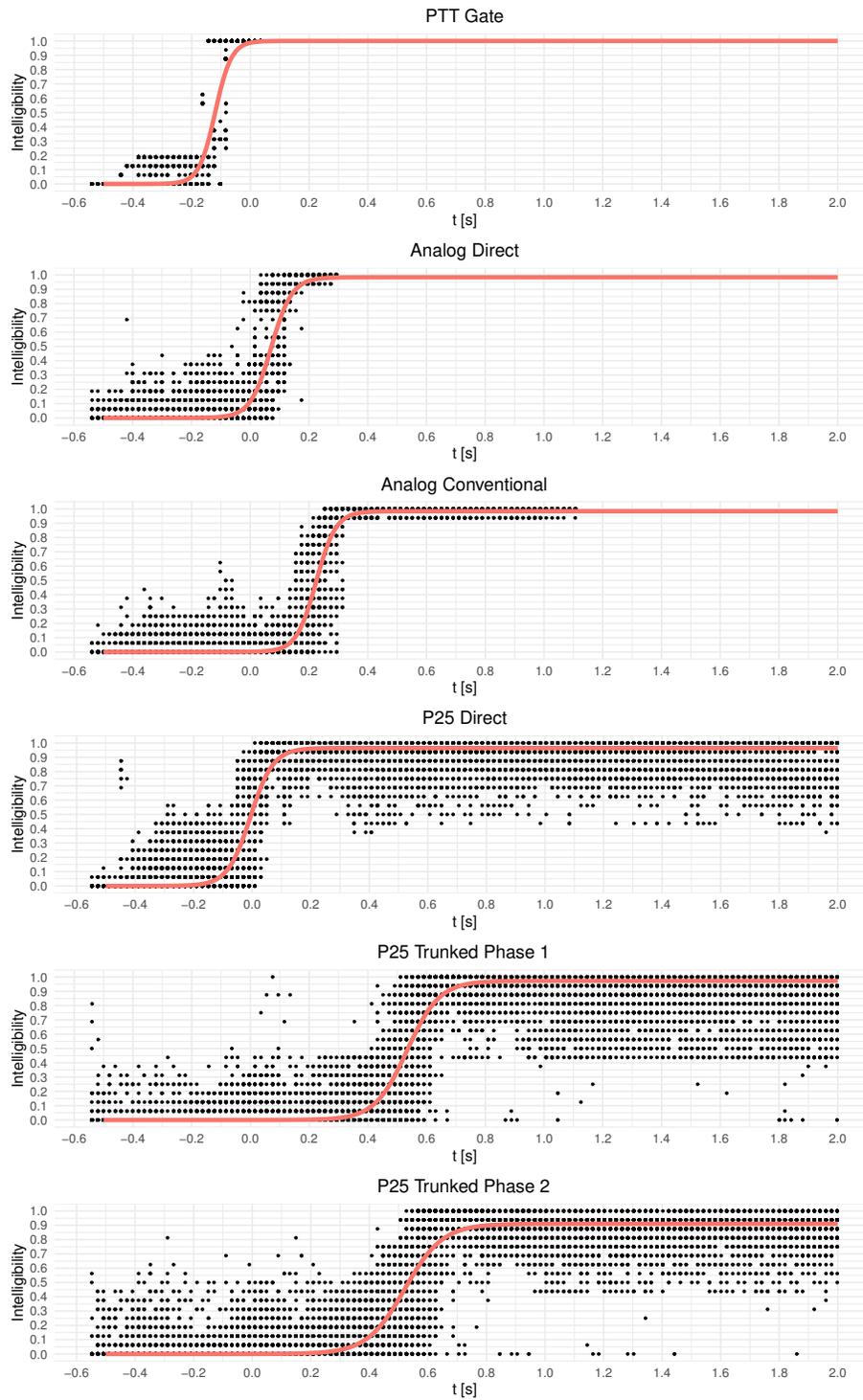
Fig. 15. Intelligibility logistic curves

request must be made by the transmitting device and then fulfilled by the system directly impacts the value of t_0. If speech goes into the transmitting device before the system has granted access, then the message will not make it to the receiving user. This causes the much larger values of t_0 that were measured. Also, the channel grant process is highly variable, which can be seen in the much lower values of λ for both trunked modes. The intelligibility of P25 trunked Phase 1 was very similar to that of P25 direct, while the intelligibility of P25 trunked Phase 2 was significantly lower. This is likely due to the half-rate P25 vocoder used for Phase 2 communications. It appeared that some combination of the lower quality audio from this vocoder along with the higher access delays impacted the ability to perform accurate M2E latency calculations. The high access delays end up providing less audio for M2E latency alignments, and the low rate vocoder appears to distort the audio enough to cause inaccurate latency calculations. Inaccurate latency calculations can severely impact the intelligibility measurements for P_2. This suggests that further work is required for accurate access delay measurements for technologies with high access delay and low rate vocoders. Ultimately, for P25 trunked Phase 1 the 90% asymptotic intelligibility access delay result was $\tau_A(0.9) = 654.4 \pm 5.6$ ms and for P25 trunked Phase 2 it was $\tau_A(0.9) = 663.0 \pm 6.1$ ms.

There are two primary ways to compare access delay curves: as a function of percentage of asymptotic intelligibility and as a function of raw intelligibility achieved. These are both displayed in Fig. 16. By comparing access delay as a function of percentage of asymptotic intelligibility achieved, the amount of time each technology needs to set up to achieve its maximum average operating mode is compared. It further allows access delay to be considered separately from the intelligibility of each technology. Comparing access delay values directly as a function of intelligibility also has functional utility. It makes it possible to select a desired intelligibility and determine how long each technology would need between a PTT push and the start of speech to achieve that intelligibility. However, infinite access delay values can be achieved if the desired intelligibility is greater than the asymptotic intelligibility for a given technology.
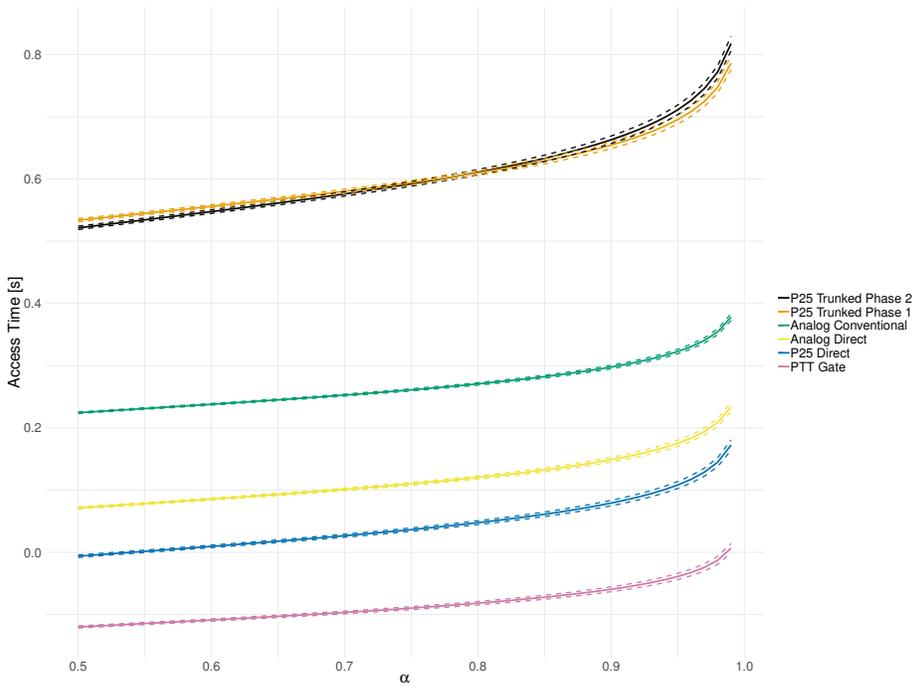
(a) The Asymptotic Intelligibility Comparison, where dotted lines represent 95% confidence intervals
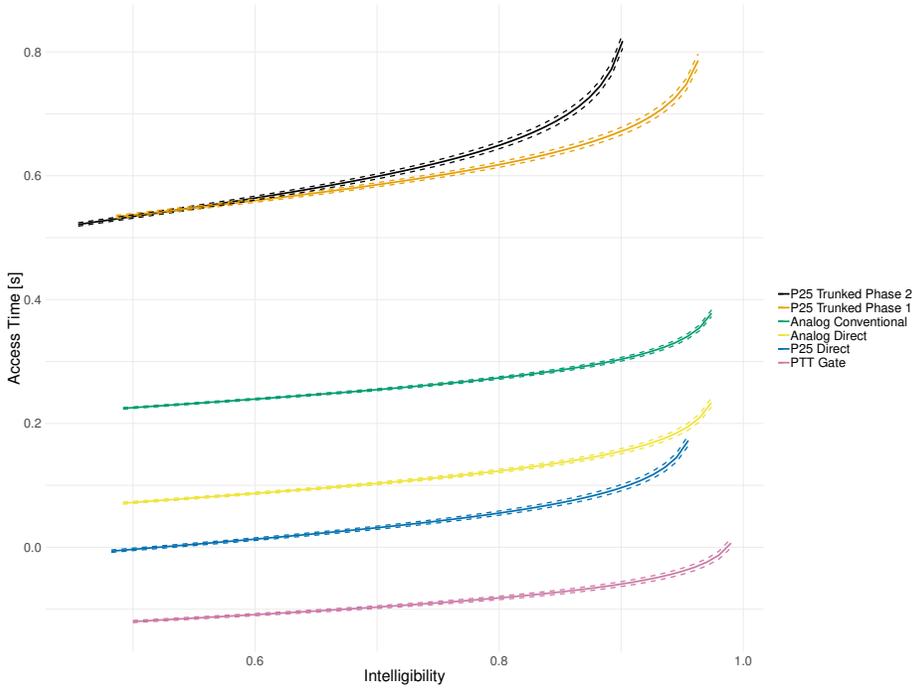
(b) The Raw Intelligibility Comparison, where dotted lines represent 95% confidence intervals

Fig. 16. Access delay results

.

6. Conclusion

The end-to-end access time measurement system developed by NIST PSCR researchers was shown to work on various communications systems. While example measurements were performed only on analog and P25 PTT systems, the design of the measurement system should work just as well on LTE PTT systems. This measurement system was developed to successfully capture end-to-end access time and is intended to allow others to perform testing on other communications systems. The focus on QoE created a challenging and worthwhile outcome: a measurement system designed around the user experience.

The outputs of the measurement system, I_0, t_0, and λ, are meaningful and easily interpreted. Together they offer information on the intelligibility of a given communications system, how long the system requires between PTT being pressed and the receiving device receiving speech at 50% of the asymptotic intelligibility level, and how quickly the system transitions from unintelligible to intelligible speech. This information can offer insight into the behavior of different components of the communications system, all grounded in their impact on the overall user experience.

6.1 Measurement Limitations

The measurement system presently requires the use of words within the ABC-MRT database. Inherently this means the system relies on ABC-MRT intelligibility results to match closely with how humans perceive partially muted words. ABC-MRT was designed with typical radio impairments in mind such as low channel quality and background noise [9].

Taking access delay measurements is time consuming, as the system depends on a large number of trials in order to accurately characterize an intelligibility curve. Since the relationship between pushing PTT and getting an access time value is not one-to-one, the tests are long and can consume a large amount of space on the test computer. Due to the length of time required to collect data, and the inherent feedback used to shorten tests, two location field testing was not feasible for this phase of system development as in Ref. [1].

Finally, the measurement system relies on accurate M2E latency calculations in order to align the audio and send the correct portions of the recorded audio through ABC-MRT (i.e., P_1 and P_2). The more audio that can be used for the M2E latency calculation for alignment, the more accurate the calculation can be. It seems that the high access delay of P25 Trunked Phase 2, combined with the lower rate vocoder, impairs the ability to make an accurate M2E latency calculation. This ultimately affects the ability to characterize the intelligibility of the system, and has a major impact on the access delay measurement. Further research is required to address this limitation of the measurement system in this difficult to measure case.

6.2 Future Work

The access delay measurements described here rely on speech intelligibility measurements from ABC-MRT. Voran [9] showed that these measurements show strong correlation (0.95)

to actual MRT results across 367 SUTs. But these SUTs were all communications systems that were fully turned on — partial transmission of the MRT keywords was not considered in Ref. [9]. Initial tests show that ABC-MRT does show the proper trend in this regard. That is, intelligibility drops consistently as more and more of the start of a keyword is muted. In order to verify that the rate of intelligibility drop is consistent with human perception, one would need to perform subjective testing using partially muted keywords from ABC-MRT. Such testing is a candidate for future work to enhance the access delay measurement system.

In addition, once the above data has been collected, it may be advantageous to use it to design an intelligibility estimator that is specifically targeted at the precise problem of partially muted words, as they appear in the access delay measurement problem. It is often the case that an estimator that targets a specific impairment can outperform (on that specific impairment) an estimator that must successfully respond to a wide range of impairments. Development of a signal processing algorithm specific to the requirements of speech intelligibility in the access delay (partially muted or distorted word) context is another strong candidate for follow-on work.

The current measurement system will be utilized to complete additional measurements. Collecting data from more active and loaded systems would be of interest to study how channel traffic could potentially impact access time. Performing measurements across both LTE and LMR-LTE interconnected systems will help to better understand what differences may exist between LMR and LTE systems and how those systems interact. Understanding how these variables impact measurements will characterize what is needed to develop a system to measure the probability of access and retention.

The current measurement system uniformly samples across PTT times throughout the test. This has the benefit of being simple to implement but may sample more than is needed. The most interesting part of the curve is where it transitions from low intelligibility to high intelligibility. As such, it would seem that a scheme could be devised to find the transition point of the curve and sample in that area more frequently. Significant time savings are possible by reducing the number of trials as even the shortest test, PTT gate, has around 750 trials for each word.

Finally, as noted above, the measurement system must be made more robust to the impacts of technologies with high access delays and lower rate vocoders.

References

[1] Frey J, Pieper J, Thompson T (2018) Mission Critical Voice QoE Mouth-to-Ear Latency Measurement Methods. NIST, Technical Report IR-8206. doi: 10.6028/NIST.IR.8206

[2] 3GPP (2017) Mission Critical Push to Talk (MCPTT). 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 22.179. Version 16.0.0 URL https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=623.

30

[3] TIA (2016) Digital C4FM/CQPSK Transceiver Measurement Methods. Telecommunications Industry Association (TIA), Standard 102.CAAA-E.

[4] Atkinson D, Catellier A (2008) Intelligibility of selected radio systems in the presence of fireground noise: test plan and results. NTIA, Technical Report TR-08-453.

[5] (2007) NFPA 1981 Standard on open-circuit self-contained breathing apparatus (SCBA) for emergency services.

[6] ANSI/ASA (2009) ANSI/ASA s3.2-2009 method for measuring the intelligibility of speech over communication systems.

[7] Voran SD, Catellier A (2015) Speech codec intelligibility testing in support of mission-critical voice applications for LTE. NTIA, Technical Report TR-15-520.

[8] Voran SD (2013) Using articulation index band correlations to objectively estimate speech intelligibility consistent with the modified rhyme test. *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE), pp 1–4.

[9] Voran SD (2017) A multiple bandwidth objective speech intelligibility estimator based on articulation index band correlations and attention. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp 5100–5104. doi: 10.1109/ICASSP.2017.7953128

[10] (2016) Modified rhyme test audio library NIST. URL https://www.nist.gov/ctl/pscr/modified-rhyme-test-audio-library.

[11] ITU (2001) Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. International Telecommunication Union (ITU), Recommendation P.862. URL https://www.itu.int/rec/T-REC-P.862.

[12] Elzhov TV, Mullen KM, Spiess AN, Bolker B (2016) *minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds*. R package version 1.2-1 URL https://CRAN.R-project.org/package=minpack.lm.

[13] Berry KJ, Johnston JE, Mielke PW Jr (2011) Permutation methods. *Wiley Interdisciplinary Reviews: Computational Statistics* 3(6):527–542. doi: 10.1002/wics.177. URL https://onlinelibrary-wiley-com.nist.idm.oclc.org/doi/full/10.1002/wics.177

Appendix

A.   Measurement System Implementation

A.1   Supplies

Download all necessary items, such as code and audio files.
Code and audio files are available at: https://doi.org/10.18434/M32085.
Microcontroller code available at: https://doi.org/10.18434/M32086
Data is available at: https://doi.org/10.18434/M32083.

Supplies List

This document is meant to assist in the setup and protocol of the access delay measurement
test setup. Please read through the full paper for additional details.

- Two PTT communications devices
- Communication system
- Controlled RF environment
- Computer
- Speaker
- Attenuators (optional)
- RF cables (optional)
- Assorted audio cables
- Transformers:
    - 1:1
    - 12:1 direct box
- MATLAB 2018a or later:
    - Audio System Toolbox (2018) / Audio Toolbox
    - Signal Processing Toolbox
- Audio Interface:
    - Compatible with MATLAB Audio System Toolbox (2018) / Audio Toolbox
    - Associated driver(s)
- R 3.5.X or later
    - RStudio (recommended)
    - metRology, ggplot2, minpack.lm, dplyr, and devtool packages

- PTT gate:
    - AQV210EH PhotoMOS switch
    - N-channel MOSFET (2N7000)
    - Resistors: 220 Ω, 560 Ω
    - Inverter (MC74VHC1GT66)
- Microcontroller:
    - MSP430F5529LP
    - Perfboard
    - 2x10 0.1 in. pitch female headers
    - Connector to audio output of audio interface
    - Connector to audio input of audio interface
    - Resistors: 330 Ω, 2.2 kΩ
    - Diode
    - Connector for radio
    - 4N35 optocoupler
- External 3-wire microphone/earphone/PTT devices
    - RCA connection
    - XLR connection
    - PTT connection to microcontroller

32

A.2 Setup

Access Delay Setup Guide

Reference Fig. 8 and Fig. 12 for test setup connections and additional information.

1. Ensure all required software is installed on the computer to be used for testing.
2. Connect a USB cable from the computer to the audio interface and from the computer to the microcontroller.
3. Connect the start signal from output two on the audio device to the start signal input on the microcontroller.
4. Connect the PTT signal from the microcontroller to input two on the audio device.
5. Connect playback voice from output one on the audio device, through a direct box transformer, to the input on the Tx device.
6. Connect the PTT isolated output from the microcontroller to the PTT input of the Tx device.
7. Connect the audio output from the Rx device, through a 1:1 transformer, to audio input one on the audio interface.
8. Connect the Tx and Rx devices to the communication channel either over the air or cabled through attenuators. Make sure all connections are tightened.
9. Turn on both communications devices to the same volume settings and set to the same channel.
10. Because of the threshold in the microcontroller, the channel two line output of the audio device needs to be run at full volume in order for the start tone to be detected.
11. If a trunked or conventional test is being performed, turn on the repeater.

Volume Settings Procedure

The volume setting procedure given here should be followed prior to any testing and should be done for every pair of technologies being tested individually, as V_{Tx} and V_{Rx} are dependent on technology. The goal of the procedure is to find the values of V_{Tx} and V_{Rx} that optimize the audio quality. Reference Sec. 4.4.2 for more information.

1. Set V_{Tx} to maximum volume.
2. Adjust V_{Rx} to volume such that no clipping is observed on the audio interface.
3. Use volume_adjust.m on values of V_{Tx} to identify maximum audio quality with respect to V_{Tx}.
4. Set V_{Tx} to the transmit volume associated with maximum quality.
5. Repeat (2-4) until the ideal V_{Tx} no longer changes significantly.

33

A.3   Performing Measurements

> ***Access Delay Measurement***
>
> This document is meant to assist in the setup and protocol of the access delay measurement test setup. Please read through the full paper for additional details. See code documentation for additional information while performing tests.
>
> 1. Set the volume settings on the audio interface and in the audio interface control panel to settings found using the Volume Settings Procedure. Check that buffer settings are ideal.
> 2. Determine input parameters.
> 3. Begin test. Enter preliminary test information.
> 4. Check battery levels and audio throughout.
> 5. See test code documentation for more information.

A.4   Troubleshooting

> ***Troubleshooting Tips***
>
> 1. If over/under runs occur, check the buffer settings.
> 2. Check LEDs on microcontroller to ensure power.
> 3. Make sure all connectors and cables are connected tightly.
> 4. Verify both communications devices are on the same channel.
> 5. Check that the power supply for the repeater is on (if needed).
> 6. Check volume on audio interface, the audio should not show signs of clipping.
> 7. Ensure that the computer has ample space to run tests. 20 GB should be plenty.

B.   Additions to the Mouth-to-ear Latency Measurement System

This section is intended to clarify the necessary additions to the M2E latency measurement system for access delay measurements.

The access delay measurement system is similar to the one used for M2E latency [1] with a few additions to allow for pressing PTT in the middle of an audio clip. The new microcontroller code is backwards compatible with the M2E MATLAB code.

For the M2E latency [1] measurement system, the purpose of the microcontroller was to make the test more like a normal use case. For access delay measurements, keying the radio is mandatory because what happens right after the radio is keyed is a main point of focus. The same basic setup with the MSP430F5529 Launchpad was used with a few additions regarding timing of keying the radio. The microcontroller code and the MATLAB radioInterface class were slightly modified to allow the radio to be keyed up at a specific point in the audio clip. The PTT command to the microcontroller was modified to take an optional delay that specifies how long into the clip to wait before keying the radio. In order for the microcontroller to know when the clip starts, a short sine signal was output from the audio interface and detected by the microcontroller using a rectifier. The required

rectifier circuit is shown in Fig. 11. The circuit rectifies the incoming sine wave and uses the internal comparator of the microcontroller to detect the rising edge of the audio signal.

The measurement characterizations for the M2E measurement system were done using a straight-through cable. Characterizing the access delay measurement system required the addition of a device that granted access in a consistent manner. This device, shown in Fig. 14, consisted of an electronic switch to block the audio until the PTT was pushed.

Steps to Upgrade to Access Delay Measurement System

1. Download the latest code: https://doi.org/10.18434/M32085 and https://doi.org/10.18434/M32086.
2. Load microcontroller firmware onto microcontroller.
3. Download the access delay audio files.
4. Build the required rectifier circuit and PTT tone connection.
5. Build the PTT gate circuit. (optional)
6. Follow the steps outlined in Appendix A.2 and Appendix A.3.

C. Intelligibility Plots for Individual Words
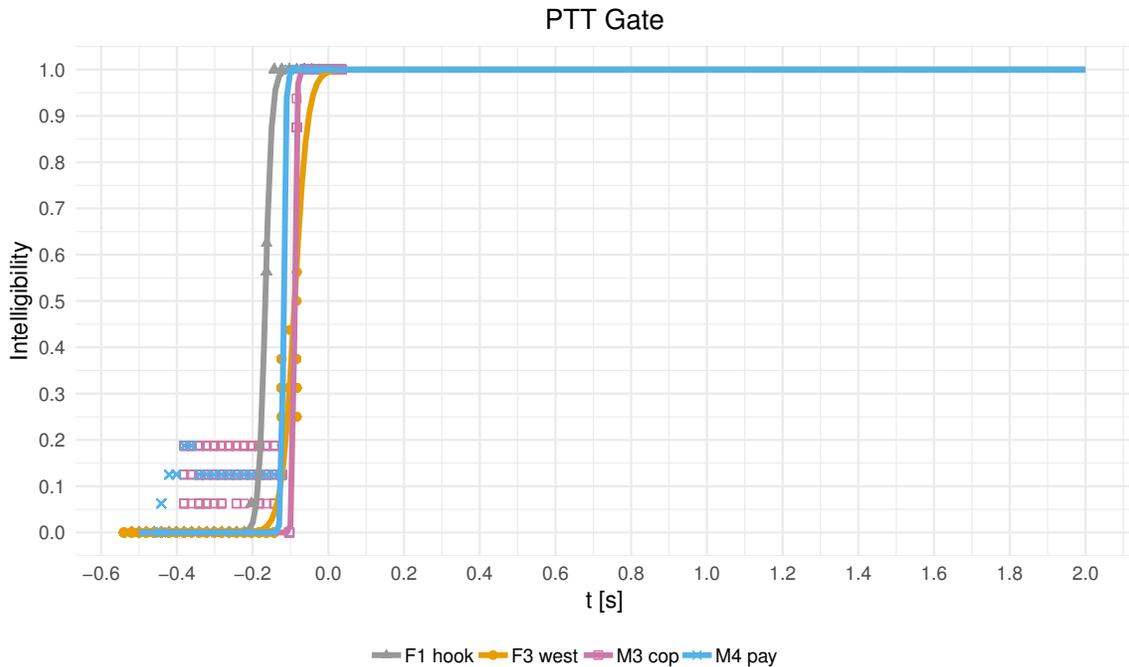

**Fig. 17.** PTT gate intelligibility by word
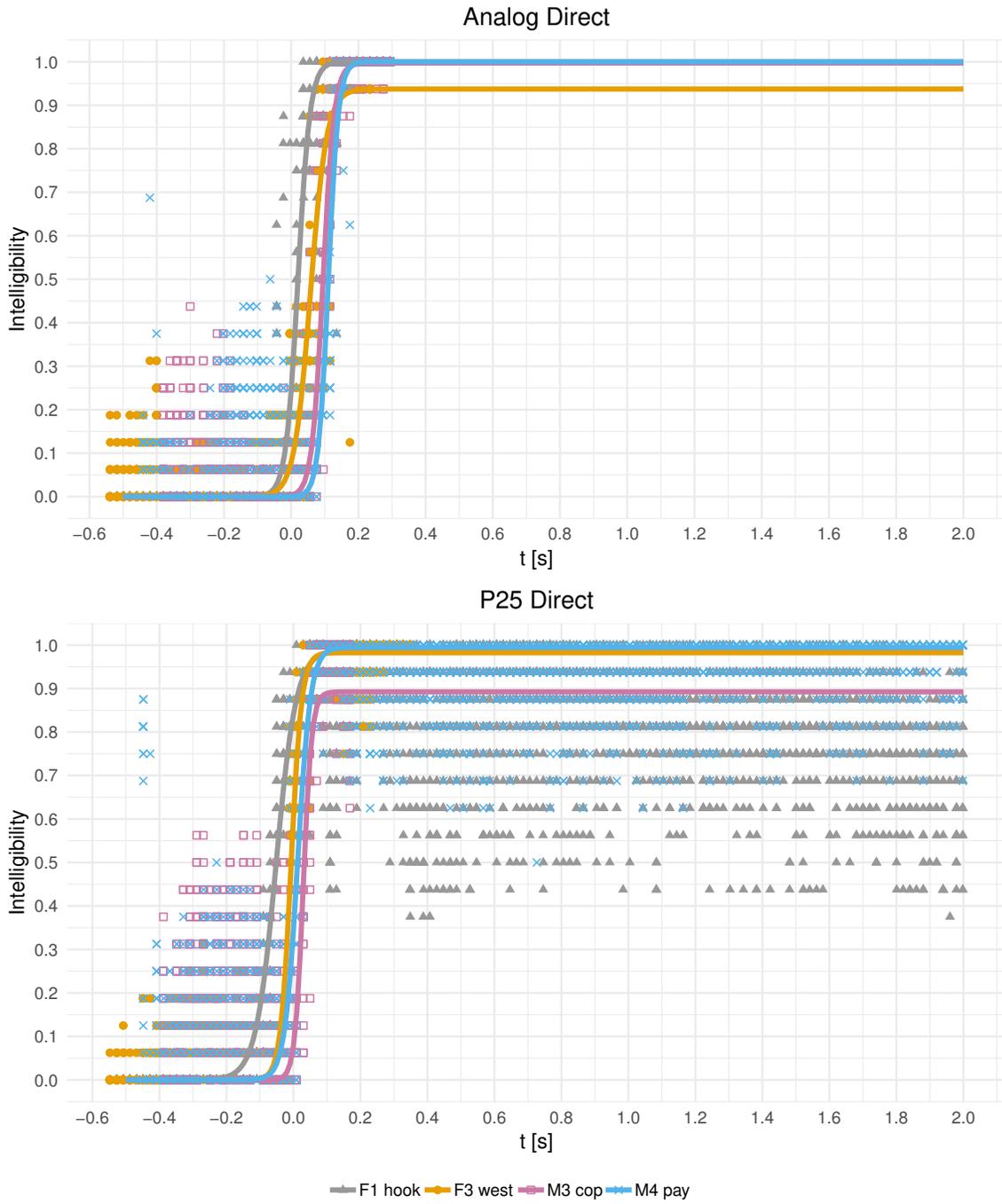
Fig. 18. Direct mode intelligibility by word
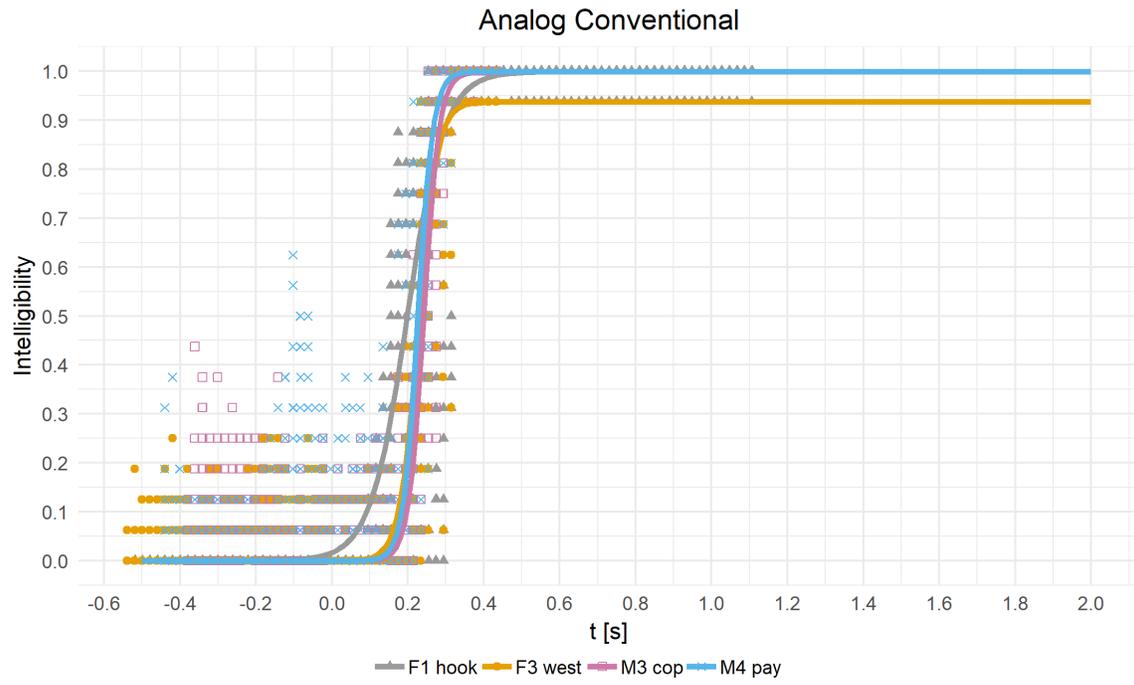
Fig. 19. Conventional mode intelligibility by word

## P25 Trunked Phase 1

P25 Trunked Phase 2
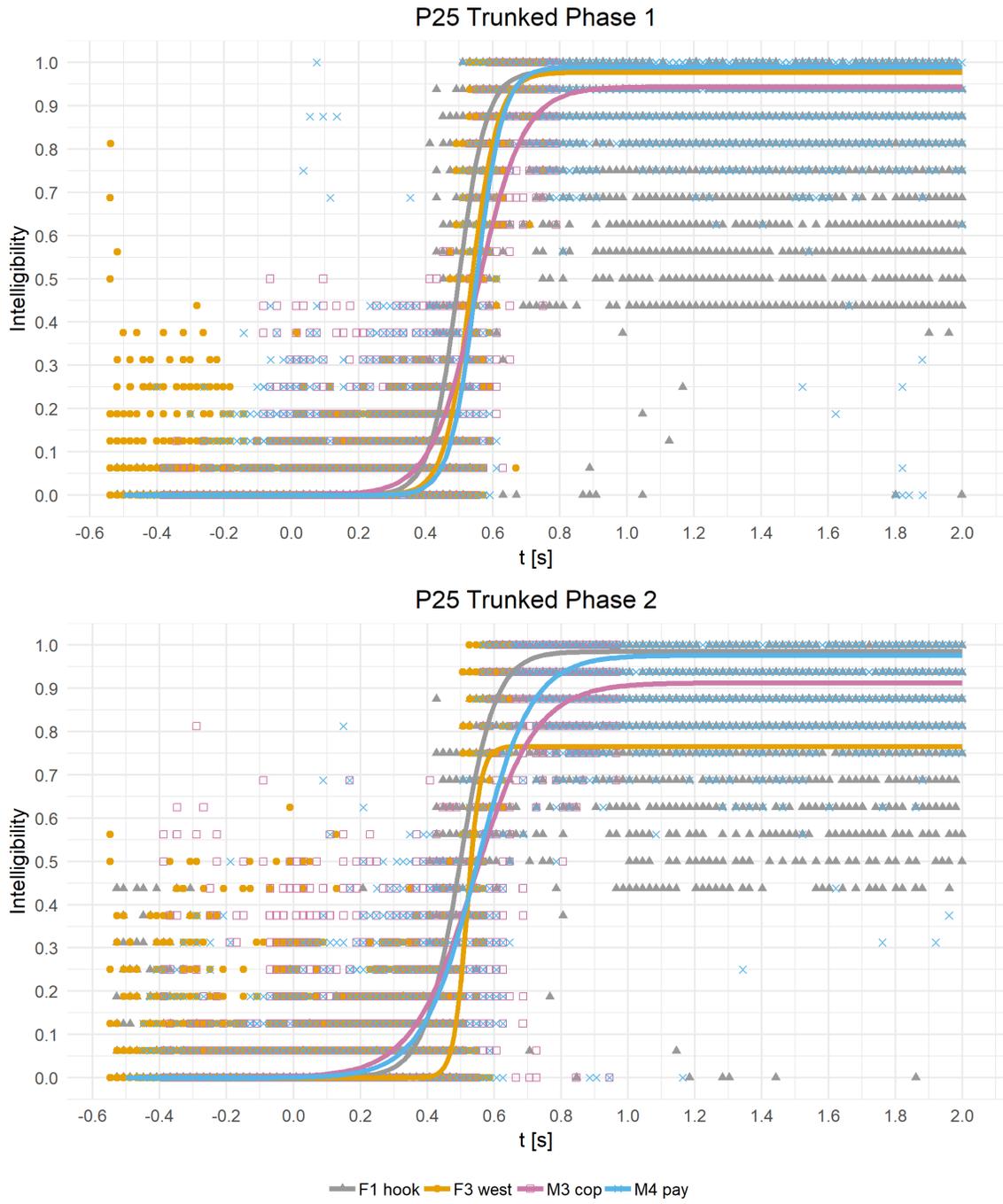
Fig. 20. Trunked mode intelligibility by word

Table 3. Example Access Delay Measurement Results. Uncertainties are reported as 95% confidence intervals

	90% Intelligibility [ms]	80% Intelligibility [ms]	70% Intelligibility [ms]	60% Intelligibility [ms]	50% Intelligibility [ms]
PTT Gate	-59.4 ± 3.9	-81.8 ± 2.8	-96.6 ± 2.2	-108.8 ± 1.9	-120.0 ± 1.8
Analog Direct	148.8 ± 3.8	120.2 ± 2.7	101.2 ± 2.2	85.6 ± 1.9	71.2 ± 1.7
Analog Conventional	297.8 ± 2.8	270.7 ± 2.1	252.7 ± 1.6	237.9 ± 1.4	224.4 ± 1.3
P25 Direct	79.1 ± 4.4	47.6 ± 3.2	26.7 ± 2.5	9.5 ± 2.1	-6.2 ± 2.0
P25 Trunked Phase 1	654.4 ± 5.6	609.9 ± 4.1	580.2 ± 3.2	556.0 ± 2.7	533.7 ± 2.6
P25 Trunked Phase 2	663.0 ± 6.1	610.7 ± 4.4	576.0 ± 3.5	547.5 ± 2.9	521.4 ± 2.7

D.   Example Measurement Results

D.1   Fractional Asymptotic Intelligibility Comparisons

Table 4. End-to-end Access Time Results. Results for 90% of asymptotic intelligibilities for each technology. Uncertainties are reported as 95% confidence intervals.

	Intelligibility	M2E Latency [ms]	Access Delay [ms]	End-to-End Access Time [ms]
Analog Direct	0.885	76.5 ± 0.3	148.8 ± 3.8	225.4 ± 3.8
Analog Conventional	0.885	78.5 ± 0.3	297.8 ± 2.8	376.3 ± 2.9
P25 Direct	0.868	220.9 ± 0.3	79.1 ± 4.4	300.0 ± 4.4
P25 Trunked Phase 1	0.875	356.6 ± 3.8	654.4 ± 5.6	1011.0 ± 6.8
P25 Trunked Phase 2	0.819	575.9 ± 8.1	663.0 ± 6.1	1238.9 ± 10.1

Table 5. End-to-end Access Time Results. Results for 80% of asymptotic intelligibilities for each technology. Uncertainties are reported as 95% confidence intervals.

	Intelligibility	M2E Latency [ms]	Access Delay [ms]	End-to-End Access Time [ms]
Analog Direct	0.787	76.5 ± 0.3	120.2 ± 2.7	196.7 ± 2.8
Analog Conventional	0.787	78.5 ± 0.3	270.7 ± 2.1	349.2 ± 2.1
P25 Direct	0.772	220.9 ± 0.3	47.6 ± 3.2	268.5 ± 3.2
P25 Trunked Phase 1	0.778	356.6 ± 3.8	609.9 ± 4.1	966.5 ± 5.6
P25 Trunked Phase 2	0.728	575.9 ± 8.1	610.7 ± 4.4	1186.6 ± 9.2

Table 6. End-to-end Access Time Results. Results for 70% of asymptotic intelligibilities for each technology. Uncertainties are reported as 95% confidence intervals.

	Intelligibility	M2E Latency [ms]	Access Delay [ms]	End-to-End Access Time [ms]
Analog Direct	0.689	76.5 ± 0.3	101.2 ± 2.2	177.7 ± 2.2
Analog Conventional	0.689	78.5 ± 0.3	252.7 ± 1.6	331.2 ± 1.7
P25 Direct	0.675	220.9 ± 0.3	26.7 ± 2.5	247.6 ± 2.5
P25 Trunked Phase 1	0.681	356.6 ± 3.8	580.2 ± 3.2	936.8 ± 5.0
P25 Trunked Phase 2	0.637	575.9 ± 8.1	576.0 ± 3.5	1151.9 ± 8.8

D.2 Flat Intelligibility Comparisons

Table 7. End-to-end Access Time Flat Intelligibility Comparison. Results compare access time values for devices at a fixed intelligibility of 85%, rather than comparing them at fractions of their asymptotic performance. Uncertainties are reported as 95% confidence intervals.

	M2E Latency [ms]	Access Delay [ms]	End-to-End Access Time [ms]
Analog Direct	76.5 ± 0.3	136.5 ± 3.3	213.1 ± 3.3
Analog Conventional	78.5 ± 0.3	286.1 ± 2.5	364.7 ± 2.5
P25 Direct	220.9 ± 0.3	71.6 ± 4.1	292.4 ± 4.1
P25 Trunked Phase 1	356.6 ± 3.8	640.1 ± 5.1	996.7 ± 6.3
P25 Trunked Phase 2	575.9 ± 8.1	692.2 ± 7.1	1268.1 ± 10.7